(12) United States Patent
Tomlins et al.

(10) Patent No.:        US 8,211,645 B2
(45) Date of Patent:    Jul. 3, 2012

(54) RECURRENT GENE FUSIONS IN PROSTATE CANCER

(75) Inventors: Scott Tomlins, Ann Arbor, MI (US); Daniel Rhodes, Ann Arbor, MI (US); Arul Chinnaiyan, Ann Arbor, MI (US); Rohit Mehra, Ann Arbor, MI (US); Mark A. Rubin, New York, NY (US); Xiao-Wei Sun, New York, NY (US); Sven Pernerr, Ellwagen (DE); Charles Lee, Marlborough, MA (US); Francesca Demichielis, New York, NY (US).

(73) Assignees: The Regents of the University of Michigan, Ann Arbor, MI (US); The Brigham and Women’s Hospital, Boston, MA (US).

(* ) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: 12/650,164
(22) Filed: Dec. 30, 2009
(65) Prior Publication Data

Related U.S. Application Data
(62) Division of application No. 11/519,397, filed on Sep. 12, 2006, now Pat. No. 7,718,369.

(60) Provisional application No. 60/716,436, filed on Sep. 12, 2005, provisional application No. 60/779,041, filed on Mar. 3, 2006, provisional application No. 60/730,358, filed on Oct. 27, 2005, provisional application No. 60/795,590, filed on Apr. 28, 2006.

(51) Int. Cl.
C12Q 1/68 (2006.01)
C12P 19/34 (2006.01)
C07H 21/04 (2006.01)

(52) U.S. Cl. .... 435/6,12; 435/91.2; 435/810; 536/22.1; 536/24.3

(58) Field of Classification Search ................. 536/22.1, 536/24.3; 435/6.12, 91.2, 810
See application file for complete search history.

(56) References Cited
U.S. PATENT DOCUMENTS
4,109,496 A 8/1978 Allemann et al.
4,683,193 A 7/1987 Mullis et al.
4,800,159 A 1/1989 Mullis et al.
4,965,188 A 10/1990 Mullis et al.
4,968,103 A 11/1990 Mc Nab et al.
5,130,238 A 7/1992 Malek
5,225,326 A 7/1993 Bresser
5,270,184 A 12/1993 Walker et al.
5,455,166 A 10/1995 Walker et al.
5,480,784 A 1/1996 Kacian et al.
5,545,524 A 8/1996 Trent
5,814,447 A 9/1998 Ishiguro
5,824,518 A 10/1998 Kacian et al.
5,856,125 A 1/1999 Mavodhalassis et al.
5,925,317 A 7/1999 Tyagi et al.
5,928,862 A 7/1999 Morrison et al.
6,034,218 A 3/2000 Reed et al.
6,080,912 A 6/2000 Bremel et al.
6,121,489 A 9/2000 Dorner
6,150,097 A 11/2000 Tyagi et al.
6,166,194 A 12/2000 Wong
6,262,245 B1 7/2001 Xu et al.
6,303,365 B1 10/2001 Wittwer et al.
6,350,448 B1 2/2002 Bandman et al.
6,395,278 B1 5/2002 Xu et al.
6,444,419 B1 9/2002 Wong
6,541,205 B1 4/2003 Yokoyama et al.
6,573,043 B1 6/2003 Cohen
6,872,811 B1 3/2005 MacBeth et al.
6,902,802 B1 6/2005 Salceda et al.
7,037,667 B1 5/2006 Affar et al.
7,125,069 B1 10/2006 Bartz et al.

FOREIGN PATENT DOCUMENTS

OTHER PUBLICATIONS

Primary Examiner — Kenneth R. Horlick
Assistant Examiner — Joyce Tung
(74) Attorney, Agent, or Firm — Casimir Jones, S.C.

(57) ABSTRACT
Recurrent gene fusions of androgen regulated genes and ETS family member genes in prostate cancer are described. Compositions and methods having utility in prostate cancer diagnosis, research, and therapy are also provided.

17 Claims, 86 Drawing Sheets
U.S. PATENT DOCUMENTS

7,638,278 B2 12/2009 Pollack
2005/0090890 A1 1/2005 Salcedo et al.

FOREIGN PATENT DOCUMENTS

WO 9902942 A2 12/1999
WO 9965929 A1 12/1999
WO 0000655 A1 1/2000
WO 000149 A1 1/2000
WO 0115386 A1 7/2001
WO 0157058 A1 8/2001
WO 016800 A2 8/2001
WO 018812 A2 11/2001
WO 0309814 A2 2/2003
WO 03051223 A2 7/2003
WO 2004092397 A2 10/2004
WO 2004097358 A2 11/2004
WO 2004113517 A2 12/2004
WO 2005007383 A2 1/2005
WO 200503387 A2 5/2005
WO 2005113816 A2 12/2005
WO 2009003423 A1 1/2009

OTHER PUBLICATIONS


Guatelli et al. (1990), "Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication," Proc Natl Acad Sci USA 87: 1874-1878.


Affymetrix NETAFXX Details for MG-U7 AV2 Microarray Specifically Showing that ETV1 is present on the array accessed from WWW.AFFMYMETRIX.COM on May 1, 2009.

Affymetrix NETAFXX Details for MG-U7 AV2 Microarray Specifically Showing that ERG is present on the array accessed from WWW.AFFMYMETRIX.COM on May 1, 2009.

Affymetrix NETAAFXX Details for MG-U7 AV2 Accessed from WWW.AFFMYMETRIX.COM on Aug. 18, 2008.


Database Entrez Nucleotide (Online) Sep. 21, 2008, “Homo sapiens solute carrier family 45, member 3 (SLC45A3), mRNA; version NM_033102.2 GI: 93277086” Accession No. NM033102.2 Nucleotides 1525-1563.


Kong et al. “Consistent Detection of TLS/FUS-ERG Chimeric Transcripts in Acute Myeloid Leukemia With t(16;21) (p11;q22) and Identification of a Novel Transcript” Blood 1997, 90:1192.


European Office Action Issued Jun. 27, 2011, Application No. 07864115.6, 6 Pages.


Jeon In-Sang et al., “A variant Ewing’s sarcoma translocation (7;22) fuses the EWS gene to the ETS gene ETV1.” Oncogene 1995, 10(6):1229-1234.


AU Office Action dated Jul. 5, 2011; Application No. 2009316693; 3 pages.


Fig 1 (CONT)

B

- Normal Epithelium
- Prostatic Intraepithelial Neoplasia
- Metastatic Prostate Carcinoma

Tomlins et al.

ETV1

ERG

C

- Normal B-cell - Tonsil
- Normal Plasma Cell - Bone Marrow
- Normal Plasma Cell - Tonsil
- Multiple Myeloma

Zhan et al.

FGFR3

CCND1
Fig 2 (cont)

C

D

E
Fig. 3

A

NPL

B

MET26

C

NPL

D

MET28

E

<table>
<thead>
<tr>
<th></th>
<th>PCA (n=13)</th>
<th>MET (n=16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETV1 amplification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMPRSS2:ETV1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERG rearrangement</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12/29</td>
<td>7/29</td>
</tr>
<tr>
<td></td>
<td>16/29</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 4

![Graph showing ERG exon 5-6 / GAPDH levels for different cell lines.]

- RWPE
- LNCAP
- PC3
- PC3+AR
- DU514
- VCAP

R1881:
- -
- +
- +
- +
- +
- +
- +
Figure 5

A

Biomarker Profile
T-statistic = 12.994
p-value = 1.4E-22

Oncogene Outlier Profile
T-statistic = 0.50
p-value = 0.59

COPA Transformation
1. center median to zero
2. scale median absolute deviation to one

B

RUNX1T1 (ETO)

C

PBX1

95th percentile score = 1.30
90th percentile score = 1.18

95th percentile score = 10.10
90th percentile score = 4.69
Figure 6

TMPRSS2:ETV1b

1 2 12

MET26-LN

142 269

TMPRSS2

71 142 365 3194

ETV1

122 220 268 322 6154

TMPRSS2:ERGb

1 2 3 11

PCA4

71 38

TMPRSS2

71 142 365 3194

ERG

37 139 225 443 3097
Figure 7

A

<table>
<thead>
<tr>
<th>Gene</th>
<th>Normal Prostate</th>
<th>Localized PCa</th>
<th>Met PCa</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETV1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETV4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETV5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELK4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EHF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELF4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETS1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELF1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GABPA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELF3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETV6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELK3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLI1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETS2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPI1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELF4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>Gene</th>
<th>Normal Prostate</th>
<th>PIN</th>
<th>Localized PCa</th>
<th>Met PCa</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETV1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETV4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLI1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELK1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPI1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETS2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELK3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GABPA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELF3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELF2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPI1B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETV6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIM2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC339321</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELF4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELF1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
AP001732 (ERG mRNA) SEQ ID NO:32

1 ccgctgggcc cgagggagtgt agtcgacccc ggctcgggcgc gcacggccaa ggacagcgcg
cctcccgcga
cctcccgcga
121 ccggcggcct aaaccttcgg ttatccagg atctttggag acccgaggaa
agccggttgg
181 accaaaaagca agacaaatga ctcacagaga aaaaagatgg cagaaacaag
ggcaactaa
241 ccgctcaggt tctgacacgc tggtagatgg gctgccctac tgaaggacat
gatccagact
301 gtcgccgacc cagcagctca tatcaag

NM_004956 (ETV1 mRNA) SEQ ID NO:33

1 gttgatagag ttccagatcc tgacggaata tcocagctaaa tgcctaaaaa ataaaaatct
61 gacgctgagtt tgcaggaaga cagcagctaa gatggatatt atgaccaagca
agtgcctac
121 atggctacca ctatgcagct tgggagaaat tgaaccgaga aacaaacaat
tgctcagaa
181 agaaacataaa ttaacagaga tcctgctcct gatccgaagag aacctttaaa
agatcctaaat
241 caattacagag aacatcgctg tgcagagaag caggtacctt acaatgagta
gcatggttga
301 ccaacactac aggtcgaagag ttggtcttta catggcctgc cacgtaaaaa
ccaagaaaga
361 ccacacagtt caagttcgaag aatcagctct gcctgacgtc aagaaacacc
ccttaaaaatc
421 agctatggag aaaaatggcct gtacaagtgc agtgcctatag atcagaagccc
acaagttggga
481 aatgcggcct ccaaccgggg cacacaccagcg aacaagccag tgcctcctcct
gccatcagca
541 tctccgaaact caccctatac accggaaaat caccggggtc cccatctcca
cctcccctca
601 tcccagttca taccagatac cagctacccc atggaccaca gatttcgcgg
ccagctctct
661 gaacctgta aaccttcatcc tctcttggcg acgtgctgaa aggaagagcg
tctgtatgac
721 cacaacccggc tggctgaaccc aacaacccc ttcacccac aagggtttaa
ccggagtcac
781 cagcaggggag tgtatgcaac caacacagag tgggagcagtc cggccagcag
aacgtttcct
841 ctccctctga tggataaaca ggaaccacga gattttgcat atggcatcaga
agtgctcgc
901 tggccacctt ctttatgcc gcaagggcgc ttccgctgtc atcacaacccg
aacagagaagc
961 tgtatatggtttaaaaaaccgggttgct ttgataagac ccttctgttga
cacagaaaaa
1021 ttcagatggag acatatcaca gagaagcagag atgtatcggg aagaccccac
ataccaaggg
Figure 9 (cont)

1081 cagggatatc accagctcttg gccagtttttg gtagcttcct tggagtagcc
ttcatact ctcttatttg ccctggacttg tcggaggactg gaatttaaac tgaatttagcc
tgaagaggtg
tggccttcct
1261 tcactccgact attactatga gaaagaatt atgcaaaaagg tggctggaga
ggatatgtc
tccatagttt cccttgtgtgacctg accagctttt gcctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgcttg ccctgctg
Figure 9 (cont)

2641 ttctgtataa ttctcttctca tgctgaatttt actatgacca tttataagca 
gtgcagtttaa
2701 ctacagagat catttcagga caaataagat gacctcaaccc attttattct 
taaaaatatag
2761 cttcgcccag gtctagctat aagcagcttt tatgcacatt gacaaatgaa 
gagtaagctt
2821 cagcctgctta aaggaacactg tgtgaacctt tgtaactttt ggtgtatagg 
aaaattatat
2881 aacaaaccgtc aaagaatagtg aggaagttgc tgtatgacat agtgcttgcca 
cagatatatat
2941 ccattctcttc tttttggaca cttctgtaaa tgtgattgga tgttttaaag 
ggaaggtttaa
3001 agtttccaaag ttttttcttc tgttttttgtc ttgcaattgg agaaatatatt 
agaagcaggg
3061 tagttgtttc cattcaccct gaaaaaacca tgtgtaaatg gggatattaga 
atcctctgaat
3121 agctcgcata aagatctacag caagggcat gatattttgtt ccatactcatca 
ataaatccca
3181 gaagaacac ttttttaaag agtctatagc aaaaaacoaa aaaaaaaaaa 
aattctaacc
3241 acaaaagtca aaataaaccct tgtaacaagc attttctgtgat gacagtaaga 
aaagattttt
3301 aagagatagct ccccccagtca cccattttccc aaaaactcac agatcaacagc 
tcatttcctct
3361 aagtggagca gttaataaga aacccaacaa ccaaaatattg tacctcttcac 
atataaccttc
3421 acaaaaaaga ctccaaatttc aaatatgta tgtaaccctgc gatttcaatg 
atgtttgttc
3481 atatactca tgtatttttt tggcccataa aaaaacaact atgccttaaa
3541 aatcagaaaaa ttttctcccc actatgcatta tgtgaggccatc tacagcaacct 
agaataaaaa
3601 cagatgttaa aatatccagt gaaaagtttta tgtgaaaaag gaattgagat 
atataattga
3661 gattttggtgct gattggaagga aaaaaatttaa tgtgaagttttt aaaaatatatt 
cgtgaagga
3721 actgtatagg ggattcattt ttgtctctttttct tttttttttttctcttatttc
3781 cttttttaata gtctagtttttt agtcagtcag tgaggaagaa ttgggccatgt 
ctaacgtatat
3841 cacaagagaa caatggcagct aatgtttaata gttataataat atttaagga 
aaactatatag
3901 ttttgcgtgtt ttaacgtagt gatcacttga actaataaca taatgacca 
acataaatgt
3961 tatttcccatt cagagagaggt tgtaaatatt acatttttaa ctttctttgaa 
aaatgtagtcct
4021 aaaatattttt aagtctcttttt ggattccacc tttttggttg agtattttatg 
ttttttttttc
4081 caagttgagtt aataaatcttg gcagctgatt tttgttaaagat ttctttgttt 
egaattcttc
4141 attgaaatgg actctcaacat aagaaatcct tttgattaatg atgtatatgct 
tttctctcagc
4201 ttttttctttc actgctgttttt gctgctctttt gattgacaca tggtaatatc 
caatagatta
Figure 9 (cont)

4261 attgcaacaa acacttatac tcaaataact aagtaaaaat aatttttctt
gttatgttca
4321 tgaaaaagtgc ttcagaataaa aatcncacaa gactacagu gcaagaacatt
cttctcaaat
4381 catgggccgga ttggcaggt ctatgttccc gtagatgtgtg taaccaatata
ccacaccttc
4441 agaatatttac aacaatttta cttatatgtc tggagggcaga agttctaaag
aagcccttaag
4501 agactaaac caagaggtatat ttgagttcgg ttcctttcgg aagctccagg
ggagactctg
4561 ccagctttca ctcttagagt ctggctgacat ttcttgcttc ctggtctcat
caactcaactc
4621 ttgtctttcga tggctacata cttctctact atagctcaaat ttcctttcctg
cctcttataa
4681 ggtagtcttgg gattacatggagggtatgct cagataatacc agggcaacct
ctccatctca
4741 agatctcttaa cttatatgacg tggccctattt cccccttggct agataattata
tctagttcctcc
4801 cagggattag cacagtgtctg taaggggtga ggccagggctg tatttcccag
acccgccagg
4861 gaggaggagag ctgggtgtgc aagaacctct aacacatttac tccaggacag
aggactttctg
4921 ctgaggtgtcctc cttgatatag aagttctata gttgcttttt atttaccttc
ggatattctt
4981 cccctcttttc tgcaataata attttgaaaa cttctctatat taatttctcctcc
ctattccactct
5041 aaaaatttctg gtaatcatac caagccctttt gattttcata atctttccccc
agcccccagg
5101 aaaaacctaa gtctatgaaac aagaaaaacag aaggtatgat aataatagta
atataccttta
5161 aatcatggtgt ctaatccaga ttatattttt taataacattt ctttgggtgt
aatatatgct
5221 tcaatgttgta ttctatcatt tgctagttgt tattacttat taggtaagaa
caatgtgtta
5281 aataagtctctt ctactcggga aaaaacttggc aaaaaatgct acacattttc
tatataacca
5341 gggagaaat atattgcaga aagatcacaaga ttttggcagag atgataggga
ttttcttaaat
5401 cagcccaacct gctctatcttg cagctttttc aagagttgta atgagaaacc
attacaggagg
5461 agaaggtctat ttggtagtttt gttaacttgga aatccataaa acaaaaaaaaa
aatattaaa
5521 ataagaagttg agtagacatttt tttccatttg cgaattttgta tgggagaagag
aggaataaga
5581 attattaaa aataacacaa tgggttaaaa tagtgggtgga aaaaaatataa
agaagggcagaa
5641 tgtatcaattt aagcaattct actaagaatt ggaaaaatca aagtttccaa
agatggtaatt
5701 agttgggcag tatactagaa aatccacccag ttttatttca gacgtcact
agatattttta
5761 ggaccccttcttttatatac atgagactctta cttgacata cttaaaaaaa
aacaagtttta
5821 tggaaagttac aagttaagag gagaatggaa ttagactaag tggatatcttt
tatagaaaaata
NM_005238 (ETS1 mRNA) SEQ ID NO: 34

1 cgccgcaggg ccggcgagga ggacgctggg cgccgcaggg ccggctggga ccggcgaggc
61 ccacccctgc ccggagctgc ccaccggccgc ccggctggga ccggcgagcc
gtcggccgcg
121 cgccgaggg ccggcgagga ggacgctggg cgccgcaggg ccggctggga ccggcgagcc
ttcocccctgg
181 ccacccctgc ccggagctgc ccaccggccgc ccggctggga ccggcgagcc
gtcggccgcg
241 ccacccctgc ccggagctgc ccaccggccgc ccggctggga ccggcgagcc
ttcocccctgg
301 ccacccctgc ccggagctgc ccaccggccgc ccggctggga ccggcgagcc
ttcocccctgg
361 ccacccctgc ccggagctgc ccaccggccgc ccggctggga ccggcgagcc
ttcocccctgg
421 ccacccctgc ccggagctgc ccaccggccgc ccggctggga ccggcgagcc
ttcocccctgg
481 cgccgcaggg ccggcgagga ggacgctggg cgccgcaggg ccggctggga ccggcgagcc
gtcggccgcg
541 cgccgcaggg ccggcgagga ggacgctggg cgccgcaggg ccggctggga ccggcgagcc
gtcggccgcg
601 cgccgcaggg ccggcgagga ggacgctggg cgccgcaggg ccggctggga ccggcgagcc
gtcggccgcg
661 cgccgcaggg ccggcgagga ggacgctggg cgccgcaggg ccggctggga ccggcgagcc
gtcggccgcg
721 cgccgcaggg ccggcgagga ggacgctggg cgccgcaggg ccggctggga ccggcgagcc
gtcggccgcg
781 cgccgcaggg ccggcgagga ggacgctggg cgccgcaggg ccggctggga ccggcgagcc
gtcggccgcg
841 cgccgcaggg ccggcgagga ggacgctggg cgccgcaggg ccggctggga ccggcgagcc
gtcggccgcg
901 cgccgcaggg ccggcgagga ggacgctggg cgccgcaggg ccggctggga ccggcgagcc
gtcggccgcg
961 cgccgcaggg ccggcgagga ggacgctggg cgccgcaggg ccggctggga ccggcgagcc
gtcggccgcg
1021 cgccgcaggg ccggcgagga ggacgctggg cgccgcaggg ccggctggga ccggcgagcc
gtcggccgcg
1081 cgccgcaggg ccggcgagga ggacgctggg cgccgcaggg ccggctggga ccggcgagcc
gtcggccgcg
1141 cgccgcaggg ccggcgagga ggacgctggg cgccgcaggg ccggctggga ccggcgagcc
gtcggccgcg
Figure 9 (cont)

1201 acggtgctga cctcaataag gacaagcttg tcatctctgc tgctgcctta
gctggtctaca
1261 caggccatgg accaatccag ctatggcagc tttcttcgga attactcaact
gataaatcct
1321 gtcatgtcttc tatcagctgg acaggagatg gctgagaatt caaactttct
gacccagtgc
1381 aggtggccag gagatgggga aagagaaaa acaaaacttaa gatgaattat
gagaactgta
1441 gcgcgtggcct acgcctctat taacgaaaaa acatcatcaaa caagacagcg
ggggaaacgct
1501 acgtgaaccc cttttgtgtgc gacccctgaga gctgctgggg gtacccccct
gaggagtcgc
1561 acgcattgct ggacgtcaag ccagatgcag acgagtgtag gcactgaagg
ggctggggaaa
1621 accctgtgca gaccttcoca ggacagcctg gttgttggga ctctgaattt
tgaattgtta
1681 ttctatctttt atatatatcc aacctcatttt ttacctctcag ggggtgaggct
 taagtcaagtt
1741 gcagctgtata tcataattgtgc ggcaatggga aaggaagacc aagactttgtg
ggggtgggtgg
1801 gaccagaaat tcttggagcca attttctagga gagggaagag gcctctctca
gaagctgtaa
1861 ggctttgccg taacagagaa agagactaat gtgtcacaac attttttaaa
atcatccatg
1921 aaaaagtgtc ttgaggttggt gcaccattag caagtgacct ttgcatcaata
gaacctatga
1981 aactgatgta aggcaattaa tttgctctctg tttttagtgc tgggagggca
aaaaagaggt
2041 gggtggtgatg aacatgatttt tgggggggggaga gcactgaaaa atctgagaac
tatttaccta
2101 tcactctagt ttgtaagccaa agatggacct cagtggggga ggctccccaaaac
cgtttgtgtg
2161 ttaaatatta ttttatataa ttggtgcca gtaatatttttt ttctaaaaat
cgtcttacaag
2221 tcgaagttgg ttcctgattt gcataaatcat gtaaggatttt tttatatttc gcgtccgagga
2281 ttctgtcaca atgaagaaaa aactgttattta tagaccccat tggaagagca
aaacgcctctc
2341 actgagatca ggctttccccaa atccatggga cttatatataag aaggaacatt
aatgtggtatt
2401 tgggtacagg gaattatgtgt gttgtgaatgt cactatcata taaaaaattt
tagccatcc
2461 ctttcttacct ttggttatag cggatttctcg ggggtgtggag ttaagtgtga
gctaagaagc
2521 attaatgcttt tgaactgtaat gatatttgcata ctccctgttgg tggacagcac gccaaacttgg
2581 agaagtggtag ggaagcctgga aagggatatcg aagggaggaag attgaacctttg
ttttttctta
2641 gtccttatata tgggtacactag atgaacctggga ataaagctgttatgcatggg
cattacccct
2701 caggtccctaa gaaataagtc ctgaatgcat gtcgttccccaa actaacactc
tgtaattttt
2761 ttatttggtc ttatttttcca agagtccctcc attttttgcac cccctctcacc
ggcaacctctg
Figure 9 (cont)

2821 ttatccagta gagagagagt tagggctttc tgattggtga gtgaanaagt
aacttgcagc
2881 acggatctat tagagaggtt tagagcttcttg gatattagaga atgtgggga
attaagagag
2941 catttcataa aatagtgcatt tgtcttgcttt tggaagagaga tgtcagagct
ttcttttgaa
3001 gaaattaaat tagcgcggtga ggatactcag tgtgactcttg tatacaaat
gaatggcaac
3061 ggtggactcag gggctctttg cttgatggtgct tgtctctggc aaggggttaa
tgaaagttaa
3121 tcctcagaaga aggaatagag ttgaggcaca aggaacactaa ggaagggagt
tcgatggaga
3181 aaatttggttg tgtcgatcaga tttggaatag ttttatcttc ccccaaggtt
aaaatcaca
3241 catagctcaag cagagtagtac atctttctgc tggatgtgag ggggattctg
3301 ttttaggaaag aagtcaacat ctaactagtg agggaggagt cctaatcaca
ccatacttc
3361 cccttcctctta cgcttgccca gttgaaaga taggagggag ggggtggtttt
tatggacctc
3421 catgagagaga ggaagagaaat atttcaggtta agcctctcag ggcggtgccct
3481 tgtgatgagaa atgggaaagt ctaactactt tctagcataat ctttaagaa
attgattgtt
3541 atattcccctc ccagccctcc acattacagga acatagctc
tgattcctc
3601 gactggctcag atactgatga attggagact ccatatggaa acitaaacgaca
gaattacccc
3661 gctgttatta ttgttatcttt tcggatgtga ggcctttaaat gagaagctcct
aatatttta
3721 gggagtgccttc aggggttttt ggggtggaact gttggactca cattgctttc
tcttagatta
3781 tgtgatttta gttgggcaacct gaaaaaggtt tgtgtgtgtga atgtgtgcat
gtgtgtaaat
3841 ttgtgtgttg tgtgtggtgtg tgtgtgtggtg tgtgtgtgtga tttgcagaca
tgcaaaactg
3901 cagctgtaaat atattctaggt atttcttaggt agttctctcc aacattcaat
aatgggatag
3961 agtgactacca gggccgggtta tcataatttg cttgtgcgttt gcaaccaggc
ataaaatcac
4021 ttctctcaaat cttccacctgt tctatattaa tttatgcccc aaactctcct
ctgctgatga
4081 ttaactcctgc agtctttata gcagataaga tataagaagaa tcgcctctag
tgcagactcg
4141 ccctgcttgt tgtcaaaaaat cccctttctct ctaagtccac cattttcaag
atggtagatag
4201 agtgattttag taagagacac cttgctgtac tggtcagagag gttttttcttc
ttgtcctaaa
4261 gggcagagac cttccagcga aaggtgggtgg gttgtttata cactggaaat
gttgcgttta
4321 tgcatttttaa aacacacagt taactccaga ggaaggatgg gcaatctcgg
tctagcttgg
4381 tgaacaaccct attttccacc aagatgcttta acctttgttg gttttggttt
tagggttcgtg
Figure 9 (cont)

4441 agtcaccttt gttccttct ctatctcttg gagggaacct ctctacatag agcccctgtt 4501 ttgctgtgcct tgggggtattg aggtcgacct ccattacatc gagatgtcctttc cctcaaccttg 4561 gagataaaca ctctggtggtt tacagccatta acctgctcaa ccctatgtgtt gagaatataca 4621 ccatctctct ctctgtcatg ctgtcgatgc cgctatctct gttgggtgtct atataaatggt 4681 gttggaactct tacctacatt ccagagaagt ttcaggggcc atatataatata tgtatacata 4741 tacatataata aatatatat attaaatat aaattacagg aataactgct cagttattga 4801 acctttttttt taaaagaact tttttttttta agctgagaag tatagggagt aaaaagaggtg 4861 tatattgtgt tgtgtactatt tccaacctgt atttatcata atttatattttttaaaacc 4921 tgggagataa gaaaacttagt gaagaaaaag aaaacagggt gctttttttaaa aactcagact 4981 gagatagctt aagatgttag cgatgtaaat gtctcatgttt ttttttaaaa aaatgcaaaa 5041 aatatctttatt ggcgaggattt tttgtttgtt tatatttagta gctgatgtgtg gcacatctatt 5101 tgtctggggaga ttttttttata tactgtagcc tgatccata ttgtattttta aacttgtgtga 5161 aataaaaaag aaaagatcttc attcataaaa aaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa

NM_005239 (ETS2 mRNA) SEQ ID NO:35

1 gccccggttac ttacccacaga gacagcagag tgcgggtgctg ctccagctca gacgtcccgg 61 agccgccccg ccagctgtccg gcctcccctga tgctctcttg ccggcgctct cgccccctggcc 121 cgccgccccg ccagcagcgg ccggcgcggg gcccggacacc tccccgacaa gcccgcggccc 181 tcgccgctccg ccggaggtag aatgattttcct aatcaagaat atggaccagg cagcctcgtgt 241 ggtcatagct tacagagggga caaccaacgg ccacgccacgg ctctgacactt tttctggtgc 301 ctggttctgct tttttttcttt ctctaaatga agagcaaaca ctctcaagag tgcccaacaggg 361 ctgtgattccc atttctcatg acctccgccaa cttgtgaattg cttttttaaa cccgtgtcagc 421 caaggtgttg atgagtcaag cctttaaaaggc tacctctcagt ggcctcaaaa aggaacagcgg 481 gcgcctgggc atttccaaaga aacccctgtcg gtcgctgtgag ccaccaggtat ggcagcttgt 541 tcctgtgccg ccacatagct tcagctctgtt gaagctgaat cttcgagaggt tccggcatgaa 601 tcggccagatg cttctgttaacc ttcgaaaaag gacgctttctt gcagctggcac cttgactttgt 661 ggtgacatt cttctgggac atctgagaca aatgatccaa gaaaaaccaag aaaaaaaaaaaga
Figure 9 (cont)

721 agatcaatat gaagaaaaatg cacacctccac ttcctgctct cattggatta
acagcaatac
781 atttaggtttg ggcacagagc agggcccccta tgggaatgcag aacacagatt
accccaaaaag
841 cggtctctcttg gacagcatgtg gtcgctgcttc cacacacacgc gaatcagcatc
tcagcagaga
901 gtctggattttg tcggggctcag gtcgctcag gcacactact
gctgtgctcag
961 tcagcatcctg ccagccagca acttgaatttt gctcaccacaacttcgtggga
ctccttcagga
1021 ccagcaacgtt ctcggagacgc gtcgagagac ttcggagagc tcagactccc
tctctcagtc
1081 ctggaacaacgc cagctgctctg tcgtggtgtg gcaacggtttt ccttcctcctc
tagagcttcga
1141 agatgaacctgac ggccagcttc ttcgcttcsga taagcccacactgtcttctca
aggatacatg
1201 ccaagacaggag agtgacccagg tggagcaaggg ctaaaccagtt atacactgcag
cctgtgcggc
1261 cgctctcaca ggaatgctgac ctatattcagcct tggggagtttt ctcctgagc
tgtatcagca
1321 caatctctctg cagcttacatg ttcgctggag tggagacaggg tgggtatgta
agctgcggca
1381 ccggcgagttg ccctgccggg cgggagggga ggggaaggggta acacacggaga
tgaactaca
1441 gagaagctgac cggggtcttc gctacttacta cgacacaagagataatcctcaca
agagctcggg
1501 ggaagctcgtc tcgtactcgc ccgggtgtgga cctgacttcga tcgggtggtt
ctcggccgca
1561 ggaagctcgtc gccatctggt gctcgcagcc ccgcagcgag gactgaggtc
gccgggacca
1621 cccctagcgg gcccagcgttg gtcggactgta ggacgccacactctcggc
agctgcggc
1681 agggagccaggg aaagggcagga ttaggaatgt tcagaggaatg gccaagacag
cagtggacct
1741 attgtcaccgg aacaccagcgc tctggaccag gtcgctctccc tttggtgagc
aacgggacag
1801 ctaatttctactc tcaagtctgt ttaaaggtga aatggcgag caagaagccac
caggaagcgg
1861 tctctggcggc tggcagtcgc gggagcggga tgttctgtgg cgttggatgat
ttcctaaagga
1921 gccagacagtt ggttgcaacac cacagactat ttttagacac tctttgctctc
acacacatgtg
1981 ggacacgcaat gccaaataac tctttagagag ggtaagaggg tgaggaagga
acaacacatgtg
2041 catttccagaa gttgtttttgt atatatatttt atatattttattttgttctc
gaatcctctt
2101 aacagaggtct ttaacacagaa atggatatttt gtaatatttataa ataatatataaatgtttattt
aatccttatt
2161 gaaataaggaatctagcgctctg aggtttcttctc ttttcattttc caatagcaacatgtagtt
aatatatttc
2221 tagcacaagatt taatctgcac cgggccagact aagagaagtt gtaaagtagtt
aatatatttctc
2281 attataattcatctcagggga taaggctgtg ggagggtaaatt ccctgttttttgttgtttttt
Figure 9 (cont)

2341 tgtttgtttg tttttttgt tttggggtt gttctgcct tgggtgtcct
gcaagacctt
2401 tgtttacattgc ggagttttta tgagaacactt aatgtttatt atctggtctt
tagactggcc
2461 ttggctttct cttcttaattc taagtaaaac gctataaaagc aagtatttttc
ttgacaaatgt
2521 gcatatgttatt tcaacctttct tgcagtgcctt taagtcgctt tataacaacaa
tagattttta
2581 ttggattgatt tataaattgt ttcggagcac acctacctctt tctgaccaacc
cagccat.ttcc
2641 ctctctgtgc tccacgttcct tctgtgtgat taaaataaga atattatatttt
tggaaataattg
2701 caacctcatttt tcaagagatca ggagggattt atgtagacgc tattttattact
gaaaaagtaa
2761 ttcacttgaa aaaaaatgta atttgaaga aagctttatttt ttcctctcag
cctctgttaa
2821 agtttaagtatt aagttgcaga gctgagagac ggggggcggt aagggtcttg
tagaaactcctc
2881 tggagcaag cacagtttgct tccatcttctg ttcactctgct tgcctcaacc
atctttaagat
2941 cattgctgctt ctttttgcct cagtgcacact cagaagagct cagattctta
tttttttgga
3001 cacagactat tctggagggac agagcggggga cttaaagatgg gaaagagaaaa
gcatgcgccgac
3061 cattttctctt cagagaggtct tttctcataa atgtgctaac tgcagttttc
ttcaaaaagag
3121 cacctgagttt atgtgttattt cacgcacatta ttaatgtacc ggttaaggct
gtttttaaggt
3181 cgcacaggct tctctgttta ctggtgctgt ggctttgagac ggggtggtga
cattgttaaa
3241 gaatctcattc tgtatgaaac tgaggaatcg ggtggcgggg caagctggga
agagcaacagc
3301 cagacgtgcgt tcgctactat accccaaaga cacacattccc agatatacata
gcagcagggat
3361 gtttttttcca agagggattg attttactct tgtacatcttg tttataatat
aacagacat
3421 tgtgctgggta acatttctgct gcccaggagaa tctagtggcag ttgctcatgg
tatttgaggt
3481 tgtgaccaagtt gaaatttcca attaggtctt ggttttttac tcaaaagaag
gatgctttcag
3541 gggttcagcc taaacgttatt ggaactaca gctctttataa accattggca
tgtgtaaataa
3601 cagacttttaa gttaaaaataa tttgtaatttg ggcctttctc ctctcaataaa
taaagatattt
3661 tgggttatata aag

//
BC056150 (ELK1 mRNA) SEQ ID NO:36

1 cgctacacac aggtacccct gggatggcct gagaactcc cccagcgttgg acccatactgt
61 gacgccgttg cagctttctgc tgacgctgct gagaagccaa gggataagccc
acacactcctc
121 cttgacctcca cgggagggat ctttttttca aacagagcag ggccacggg
Figure 9 (cont)

```
181  gttggggctca  cgcaagaaca  agaccaacat  gaattaagac  aagctcaagcc  gggcttggcg
241  gtactactat  gacaagaaca  tcattccgcaaa  ggtgagccggc  cagaagttcg  tctacaagtgt
301  tgtgctctac  cctgagttcg  caggggtgctc  caagtggagac  tgcgccgcgcc
agccagaggt
361  gtctgttacc  tccacactgc  caaatgctgge  cccctgctgct  atacatgccg  cccccagggga
421  cactgtctct  ggaagcccaac  gcacacaaaaa  ggggtgcagga  atggcagggcc  cagccctggtttt
481  gcacgccagc  agcggaaacg  agtcatgctg  ctgggggcctc  tattccactt  tccacacaca
541  gttctgctgc  gcgcagccac  cccctcctcc  tggccctgct  gtgggtgtctcc  ccagtgcagc
601  tcttgcaaggc  gcagcagccgc  cccctcgggg  gacgcacaggc  accagtcaca  gccctttttga
661  gccttgtctg  ggcggtgaag  aggcgccccct  gcctctgcag  gtcctctccta  gccttctctga
ccccgcccca
721  ggccccccac  cctggaatggc  aagagcttaaa  tgtggagcccg  gttttgggccc  gggctttggcc
781  cccacaagttc  aatgtgaaag  ggcctcaagag  aqagttggga  gttggggggg
agagaggttt
841  tgtgccaagaa  accacaaaggg  cgcaccagaga  agttctcctca  caggagggcg  tgcaccgccg
901  gctgccccgccg  ttggtttattgc  acaccgccagg  gcagggccggg  ggccatgcggg  cttccagccc
961  tgagatctccc  cagccgcaaga  agggccggggaa  gccccgggac  ctaggccttc  cactcgccccc
cacctgagcc
1021  gacccctgcta  gttggggcggg  gaccccaagcc  gaccccagaag  tgggaagagtg  gctccggccct
1081  cagggctcgcg  gggccggccc  tggcccccctc  cctgcttctct  accgcatatat  tggctgcagt
1141  gctgctcggc  cccagctcgcc  tgccctctctg  gttctctatag  ctcctacttc  tggagaacccc
tgagctccat
1201  tgcgcggcgt  agcggcggccaa  aqgcctctcct  ccagtttccca  tccagttgcca  gcgcctcaggt
1261  gcacatcctct  ctctatcgcc  tggtgcccccc  tctgagcccccc  gttggtgcctt  cccccaggccc
1321  ccaggaagcc  tgcatttacac  caccaccacc  accaccctt  cttgggtctcc  tccatccatg
1381  ctctctcagc  ccagccactct  caaggagaaaa  catagttccaa  ctagaacacet  catgctctgta
1441  tgtgtgtggtgg  gttgggtcctc  tggagagaaag  ttacctcttccca  qagtaacctc  cattatctcc
1501  tccacagaaaa  acacacagct  tccacaaactt  ctctgtttttc  tgtctagcccc  ccagttgcccc
1561  cccctcacagc  tctctactcct  caattgtaggg  ggcgggtttata  ttattatttt  ttngaagccc
1621  acctggagaga  gcctgaccca  acctttttagg  gttgggttaaaa  catctcccccc  acctccccac
1681  tttttttccc  aagacaagac  aatogaggttc  tggggtgaga  aacgaaccttc  tttctttattt
```
NM_001937 (TEL (ETV6) mRNA) SEQ ID NO:37

gctgccttg ccccctgggga aatgagggag ccctgtctgc gtttttggat
gtgaagtagaa
1801 gatggtagttt gttttgtttt attatatctg gcccatactca ggggtcagg
2080 aaggaatatgc ggtttggggag tttttgggaa aagaagatca cacccctggg
atagaaatgg
2191 ccacccccttc aacccctttct tcagacagct tttctttttc aacaaacctt
2314 ttggccaggg
ttgatatatt
2461 aactctatatt tttatattgggg aagcagagga gtcattcttt
2221 tgtgggtgggg tgtgggaaggg aaggtggtgg aagagatggc aacctgcgct
2321 aaggtgtttat ttttatactc aaacagtaat tttcgattccc ttttttttcc
2451 cggccaggg aatattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Figure 9 (cont)

601 aggtgatgtg tctctatgaac tctttcagca tattctgaag cagaggaac ctctgattcct
661 ttttttcaac ttttcccacc ctggaaaaactc tatacacaca cagccggagg tacatactgca
721 tcaggaacct gaagaagaata actgtgtccca gaggaccccc agggcatccg tggataaatgt
781 gcacccataac cccccccacca ttgaactgtt gcacccgtcct gcagtcacca tctgacaaaa
841 tccgggctct tctctgcaccc cccagccagcc gcctctctcg ccctcccccttgg acaacatgat
901 cccgccccttc tccccggcttg agagagctcca gggacceagg cccgaccagg agaacaacca
961 ccagaggtcc taccctctgt cagtgcttcc catggagaaat attcaactgccc
cagcgttcttc
1021 cgagttccac ccagagccat ccagccccccgc gcaggagagc aacgctgtga
tcagcgtgat
1081 gccgagcccc atactgaccc ccctgatctct gaaacccccgg ccctccgttg gatttcaaaca
1141 gccagggttc ccaggaggcc gcgtgcatag ggaagggaga ccctacaacc ctctctcatcg
1201 ggaagacccct gottaacatga accacatcat ggtctctgtc tccccgcttg aagagcaacgc
1261 catgcctcatt gggagaataag cagactgtag actgtctttgg gattaactgtct atccagtgtcg
1321 ttcgtgacac gcgtaagaaa acctcactccg atggggagac aagaatccaa aaatttcccg
1381 gatagtgtgat cccacagggac tggctcgact gtagggaac ccctagaaca
gaaacaccaat
1441 gacccatag aaaaaatcaca gcacccgtgac ccacatcacta aacactaaaca ttatacagga
1501 ggcgccagga ccaaggttttt tgttccaggtt tattgaacc ccagatgaaa tcagtgattgag
1561 cccgaacacgc cgtctggagc acctcaggtc cccggagctg gatgaacaaa tataccacaag
1621 agatgaatgc tgaaggaacc aacagccccc acctagccac ctcagccgggc cagcagccca
gggacccccct
1681 gccccacagg attgctggaa gttgtgacggga gcagcgcggcc tgaggagagt ggaaaaagga
1741 ggcgcacagga aatggcaggg acacttcctt tgcaagacca gagggacttt ggagacccct
1801 agacacaacta cccagcacaag gcgggtgctgg aatttctggcg gaggccatga gcctggagacct
1861 ccattgcacgc tttctttctg attttggaatc ttttcctctg taattctcctgc ccccacccct
1921 tccacccgttg tagtatctat ggtgtttttg ttttgggttc tttttatatg aacctcaggttt
1981 tgcattcctca tcccttcctct aggggaagac atctgatgtt gttttctctat ggaaatatata
2041 atcttttatata tatatatatttt ttgcaaatct cacaatgctgc gccaagccca gcgtgtgcaggg
2101 aagagaatca ctggcagagg gtttcaggtt ccccttttct ctcgccacgtg gattcaggctt
gctggtcaggg
2161 gttctctggta ctgttggtgct ttggtcagaa aaaaaaaatg cttttaaaaa agataaaatag
Figure 9 (cont)

```
2221 aaaaagagag ctctctctttt ctctctcttt ctctgtctct ctcttgtctt ccctttgtcc
cctctgtctt
2281 ccgccctgcc ctgcaggtgga gatttacagt gcctcgacat ggtggttgat
tcttgagagag
2341 tcttgagat tgtttgcacc taaaacagat cagtgacccc ggtggtttgt
ggacacacagc
2401 acacaatacag aaccgcacatcc cacgatgggg cccacccattc gagggaggccc
aaatatcatca
2461 cagatgtctgc tgtgtgtgcag cagagatacat gctagtcaccag agagcccgc
ctggtatggc
2521 tttgagacac atgtgctctaa ggctgtaagat aagagatgag ggtggttcacat
gttttgaga
2581 aggacctggg agttggagat tagctttggga ggccagcgcgt gcacagttga
aatagaaa
2641 agacacacaca gggccacagc ttcagaaact gggccaaatgt atctctctgagt
ggggattttat
2701 ttctcctcag caggttaacag aggggtgtga gaaggagaag gctctctctg
tctcattcaca
2761 aaggcccatct tgtggtgtcag ttctatgccctc acctgattt ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
NM_016135 (TEL2 (ETV7) mRNA) SEQ ID NO:38

1 gatttcccc caccgccacc ttcagtttctt ggagccaggt taggggttttg
gcgaggagg
gggggggggg ttccggtag ccccccagca gccagggc gggggcgc gctgggggctg
cagatgagga
gtttggtggc tggcgtggtgc cccggtgctgc gatgagga
gctgctgcg tggcgtgcc ctgcctgggctg
cgtgccgtggc
gagggggtttg
gaatagggagaga
gatgagaga
gagggggtttg
gagggggtttg
gagggggtttg
NM_004454 ERM (ETV5) mRNA SEQ ID NO:39

1 gacttccagcc gctgtggtcgc ggagcgggctc accegttctcqc gacgcggctcg gcccagccttt
   61 tgcggccaggg gcggccagcgc gctgtggtcgc ggcctgcg gggccgctgc
   121 ctcgaagggc agggagcggg cgcgttcagg gggatccttt tgcggccagc
eattaccaaa
   181 tgcggccaggg cttcgaatagg gtatagagag cgcgttgagcag acatgggcag
   241 acctgcagtc ccttattag tgcggccaggg atctgtgatcc gggagtcg
gggggggccc
   301 tgggtgatgac agagaagagga gttttttgga ctcagatctg gtcagacgtt
tggaagagct
   361 atttcaggt ctcagtcac ctcagagagc tgggttactg ggaacacaag
ttcctgtgagta
   421 tgaagacgttt gttccagatt ttcagctgta taacctggtg ttcctgtgagct
cacctcaccac
   481 cagctacaaa ggagacgtcg acagccccctc ctgctgagctg tggctggtctg
   541 gctccccagtt gcatactagt gagaaaaagtgc cccctacaca tatcttgctct
atgataagga
   601 gctccctcctc cggttcagac cttacaacc ccccccaac cccctcactac
ccaccoataa
   661 gaatcccccttttccgccac ttcagggcagc tgcggccaccc tcaaggcagc
cctgccaccgc
   721 tggcccaagtt caggtttgag gcgcccgccc cggccccccc cttccacctcg
agcttgaggct
   781 aacacagccaa acatgtcggg tcccccgacc acacactcag cccctcgaga
tgccaagagat
Figure 9 (cont)

941 gatgcccgtaa aaccagtact catcagaaca gagatttcag agacaactgt
ctgaacccttg
901 ccacccccttc cctctctcagc cagggagtttcc tgtgagataat cgcccccaqtt
accatatgca
961 aatgtagagaa cctattgtccct gcgcagctcc cccgcccccc cagggatgca
aaagaagaata
1021 ccagacccc gttatgaacc atgggctccc gggcatgccg gggccccccag
cacaggaggttt
1091 ccagtcacca atgggaaacta agcaggaccc tcgggattac tgggtgcatt
cgagtcgagc
1141 taactgcagc tcatacttaca tgagaggggg ttaattttctcc agcagcccatg
aacggtccc
1201 atatgaaaaa gatccccgacata tctactttga agacacttggt gttggtgcctg
agagactcgcag
1261 agggcacaatgga aacacagggcg ctaactagata cggagagggg cccctttacc
agagggccag
1321 ttctcttcag ctgtggcagct tctttgttac cctcttttgat gacccagccca
atgcccactt
1381 cattgcctcag acaggtcaggt gcattggtatt caagctgata gaacggggag
aggttcgcgg
1441 gcggctgaggggc atccagagaagccggccacgc catgaaactat gacaagctga
gccgtcctct
1501 cgcgtattacc tattaaaggg gcattacgca gaaggtggct ggagagcgat
acgtctacaac
1561 atattgtctgt gacccaatgat ccccttcctcc cagctttcct ccggataacc
aggttcggttt
1621 cctgagggca gatccccagct gccacctcag cggaggggac acctgtgccc
tgacccactt
1681 tgaagacagc cccgctttacc ttcttgacat ggaccgctgc acgcacccctc
cctggtcgg
1741 aggtttttgtct taaataggtt gtaagttggtg cagtggtgcgg aaccctagac
ccaggttcc
1801 cattcagggca aacaacggtta gttgttttgt tttggttttt gttttttgct
aaagcttgcag
1861 ctttgaagtagt tatctcagaga acccaagctg tctctcgtatt gggaccccta
aagagagata
1921 cattggctgg ggagtggggaa cagggagggg cagaaaaacca ccaaaaggcc
agtctccaa
1981 ctctttgattc tagtgaggttt ctggggaaga gattaaaaag tggctctctt
accatggaga
2041 atacatgcaa agcaatatct tttttcaggtt agtacccggca aacggtgaca
tgatgtgcag
2101 atctccatgc aatcctggacc actaataattgc ctttacatag aagggtctgt
attgcacaa
2161 tttgtttgaaa aatcacaacac ccataagaaa tgtgagttgagc taagttgggg
aggtctcaac
2221 cattaaggggttaaattcact ttttaaactattgaaagccttctactgctga
atctgaatata
2281 ttcctcccttc tctagaaaaa gggggcagct cagaaacagct cttccccacact
cggttttccc
2341 aatcataaa accatggcta cctttgagga ccaacccggggc atgtggtcgc
cagtagagc
Figure 9 (cont)

2401 aagcccccctt tctcttcccc atcaacgttgc tgaagtgtgga tgaaccttat 
tattaggaga
2461 gggcgattaa cacccccgcc agtatatttttg ttcgccccctg cttggggtat 
tggttttttt
2521 tgtgttgtgt ttcggaaaaa cacccccctac atgttgttgg tattaaagc 
tatttttttt
2591 aaaaaaaacga aaaaaaaaaa aaaaaaaaccca accttttttggt atattgtcact 
gtgtttttag
2641 ccagggcctgt gcaaacttag aagcaactgc agccttgagag ggcttttct 
gagcttcctcc
2701 cccgacccatag ttcagctaacc cctttttttgcc ttcctgccgc cctttttttg 
agcaacccctg
2761 atggaacctt gtcactctgag tgtatactttttg tgaagctgct gaggccagcc 
ctgaccaag
2821 cttgattttcc ctaagtttag tccccccgtgt tcactaaccgt gcccctctctg 
gaaaaactaat
2881 tgtatatatcc tagttgtactgt cctttggaaa cctgtgctacc tggtgaanac 
cctttggatt
2941 cccctcctggc cagactctgg gcaattccc atttctatcac tttgtttttt 
cccccccttt
3001 cttttatctgt gacacccaa accctacccgt tgcgatccag tagggttttga 
gactttcttg
3061 tgcatactgac aggtaagcga cagtgcacgc attctagatt cctgccttttt 
ttttttttttt
3121 aattttttttc attttggttat tataattggaa aagtttttaa caaccaagct 
aaacgcattgt
3181 gaagtttgag ccctaaatgtag aggaaagtgt actggtggtgta ccttttttgc 
tgtctttttg
3241 gtgaaactttc gtgccctcccgt gacactctgag tataatagag atgactacac 
tgcttttttt
3301 atgtgaagaga ggcagtgctgt actcctgtag tggagacac gtccttttgaa 
ttcttttct
3361 attcatggag cactccataag tctccaaactg tccccccctat gaccaacaagc 
acatttggta
3421 agaggttcgc agggataagg ggtgcaccttt atagctattggaacatgaga 
tttctctcta
3481 tggggaatctc attagcccca aaggtgggtac aacgtgtaga ttgggctctta 
atagattgtt
3541 tattcataatc aaggaacctt ttctaaaaac tttttatatc ttttttttac 
tacacatatgt
3601 ctatacatag atgcatatat taccucccagc tggtgtagaga tttattttgt 
gtaaaattctg
3661 tatagatattg gttctcctttt cttactttac cctggctttgg atttttttttt 
ttttttttttt
3721 tgaaatgatt tagttgtgtct tagcaaatag acaataatcc ttcgtagcttt 
gagctacccc
3781 tccccctgtcg taaccttactg gacctgtgcc gtcactgggct ataggacagc 
gggcactcaac
3841 tgtgctcccc atttggaactca tgacccctcccc ggtatggtttt tgttttttttt 
ccccccacccg
3901 tggggttttgt ttagggtttcc ttttccagac gttgggaaag cttacaagtgc 
agaaaggctt
3961 gaacccgcca gctgatttga aatacttttca cctgtgccag gggcgtatgc 
ctctggccaa

1501 acccaagtct cagtggttgt gtctctctagg aatcgacagt tgcatcacgt aacactccaa
1561 acagtgccac tcaccaacagt tatagccagc acagatccat cagcaggtac
tggatctccag
1621 aagttttatta tacaagccat tccatcatca cagcccctagc cagtactgaa
gaaaaagtgc
1681 atgctgcagc cacaaaaaggg ggctctctct ctctcaattgg ctctgggcac
tggcaggtt
1741 cagcaggtcc ttagtagaa agttgcaatg gaacggtagc
tggtgctttcc
1801 ttctcactcct tcaagtctacc tgcacctcttg gtggacttttt ctctgctggc
ctcagatcgt
1851 gttgctcacc cactggccac tgtaatcact tcaagtataca aaactaaaga
aagaaaaact
1921 ctctcagagag aagtagagaa aaggaatctt gaagatcatt tgaagaagaa
cactgagaaa
1981 acggacgcac agccacagcc ttatgtgatg gtatgtccca gttcactatg
attttcttct
2041 caggtagcta tgaacaaaka cgaactcttg gaaccccaact ttttttagtt
aatataccaa
2101 agctttatgga taatgggatt ttaatagac attttcaatt atatgcagac
tgactgattc
2161 taagataaat ctaagggagg tttctatatttg tgaatttttg taaaaatagag
ttaatattga
2221 ctcttttagaa tggaggagga aaacctcaact gttttctcttt gtatctaaa
tttctagaa
2281 ttcaactcgtg aaggaacagg cattttacac tataagacac ttcttttagg
attttttttt
2341 cagttgctat atcataagca ttttaaagt ttcttttcta attttataatt
gtattagatt
2401 ttctgattct tttgttaata cagaacctta aataagacgc acaagaaaatt
tatatagagaa
2461 catatattcat tccacttggt taaaactagt cttgaccttt tcaaatgcaaa
aaaacattt
2521 tttgctttg ttaaaatatt ggtgtcacttt agattgactt tagttgactg
cactatataa
2581 tataagacta tgaatatgtgaaatattcg aaaaaaggg aagtggctggtg
gttggtgctga
2641 cccgttttca gaagcagagt agtataaaag cattgagcta aqataggcac
tcccaataac
2701 tagctatgta atcttgacct tttggtggct tagtttcctct cataaaagga
agagatgtat
2761 tggattagac tagattatac ccaacctcttc ttctatgtctt aatctattta
attttaaatc
2821 ctatattttcc aagttttgct aataaaatca ttatcaggtt attttctcaat
gtaagaatag
2881 cttaaatgttg gcagagaaat aagtgaacca acaaaatttt tttatctgttt
atgatagaag
2941 tctgcttcat cattttctta aataatttttc ttataaatcttt tttgacccac
tgtgctttgcc
3001 gttccatag taacctttgttg tgtctgaagt gcacaaagaa tactgtatttt
tggagagatt
3061 caagactcttt ccttaagggc caagaaagca acttgagcct tggctcaatc
tggtgagta
NM_001986 (ETV4 (E1AF))  SEQ ID NO:41

gccccgcgtcc tgaggacagg ttctggcccc cgcttggggc cccgcccgtg cgcccgagg
61 gagccgcggg atggagccga ggatgaagac cggaactcttg gaccaagcac
tgccccctac
121 ctccagcagc aacatgcccgg gaaatgggag ctggcgcggaa gcgcgtgactg
gccccgtgag
181 gaagctcatg aacccgggct ccttgccgcc cctggactct gaagatcatct
tccggagatct
241 aagttcacttc caggagaagt ggtgtcgctga agctcaggtta ccagacaagt
tagcagatgt
301 gggtttctgt tttccattcg aaaaaactagc tttccacagc ccacaaacca
ggataaaga
361 ggagccccag atccccccga cagacccggg cctgtctctgc accagagaacgc
cgccacccc
421 ctacacacat gccgagcagt gctttaactc cagttgcttac gaccccccccag
gacaaatcgc
481 catcaagtcc cttgcctcttg gtgcctcttg acagtcgccc ctacagcctc
ttcggcggc
541 agagcaacgg aatttctctg gatccctctgg acccttccagc ccaccccttg
gccatggtta
601 ctctggggaa catagctcgc ttttccagca gcctctggac atttgcccact
cctcacatc
661 tccaggaggg ggcccgggaaac cccctcccagc ccccaacaa cacamgtgt
cggacccctg
721 cccacccat ccccacagag gctttaaagca aagataccat gatcccctgt
tagaaaggacg
781 ggccgaccca ggcgtggacc aaggtggggt caatgggcc aagtacccag
ggcccggggtt
841 ggtgtcacaa caggaacaga cggactttgc ctacgactca gatgtcaccg
ggtgctgcac
901 aatgtacctc cccacaggg gctttctctgg gcctctctca ggtgacccgg
ccatsggcct
961 ttgctattag aacacctctgc gaccattccc agatgatgct cgggttggctc
cctgaaatt
1021 tgaaggagac atcaagcagg aaggggtccg tcgatattcga gagggccgacg
cctacacgcg
1081 ccggtgacct ggacagctgt ggcgttccttg ggtggtgccttg ctggatgacc
cacaatgacg
Figure 9 (cont)

1141 ccattcatt gcctggacgg gcgcggggaat ggagttcaag ctcattgagc
tcagggaggt
1201 cgccagggta tcgggctatcc aagaagaacgg gcacacagcag aattacgaca
agctgacgcc
1261 ctcgctccga tactatttag aaaaagccat catgcagaga gttggcagtgtg
agcgttaagct
1321 gtacaaggttt gtgtgtgagc cccagggcct ctctctctct gcctcccg
acaatccagc
1381 tccagctctc aaggtgtgagtt cagccgggcct tgctagtgtag gaggacagac
tccctggggc
1441 ccaactgggt gagaagccccg ccctaccttc ccagcgtggtg gggcagccccc
agccatgtgg
1501 ccggcaaggg ggcctacttt actaccccgc aggggctggtt ccacccctgcg
caggctgggtg
1561 ctcgccccgtg tacagataaa taatctctgt tgtgggggaa cctctatctg
aaccaccagc
1621 atgtctctgg gcgaagatccc caactgtctca ccaatggtcc tagcccaagac
tctgagcgtc
1681 tcacccggagt catggtgagaa gaaaaagtgga gaaatggca aatgtcagtc
tcagaaacrtc
1741 cccctggggggt tccaccccgg gcctggagga atcacgtcga gtcctgctct
aggtccacacg
1801 ccggccccac cocccccccaa ccacacagaa caagatgttg tgtcttgtctg
gggaccagag
1861 aagggcttcc caaacttctat ctcggcaagga ggggtgaggg gtctcaagtg
tcencagat
1921 cccccactgc gggggagacaag aggcttggac tctgcccccac cctgtgggcc
tgagggtac
1981 cgggtttgtca gtttctgggt gcctgtgttg ccacagggag gggaggttg
aagaaagaa
2041 cctgggtgatg ggggtgcttg gtataagcg aggggatagg gtctctgtctc
caagggccc
2101 tttgccttttc ttctggcctt tcttagggcc agggcttggttt tgtacttttcc
acctcccaac
2161 catctgcccag accttaataa agggccccac ttcctcccaaa aaaaaaaaaa aa

NM_006494 (ERF) SEQ ID NO:42

1 tctgagaggg gagggcggtt gaggcgccgga gggcggcccg aacggcgccg gacggggacgg
61 gcgcggcgg ggcgcggggg cccggcggccc gcagcgccgg ggttcggcgc
gggcggccca
121 gcatagagac cccggcggcag acaggttttg ccctccccag ttgggccctac
aagccagagt
181 cgtccccctgg ctcagggcag atccacgctgt ggcactttat cctggagctg
tcggggaagg
241 aggagtacca gggggctatt gcgttgccag gggacatcag ggaccttcgtc
atcagggcc
301 cttgctggttg ggccggcgttt gcagttgcgg gcccggaggt aacgagacg
aattagcga
361 agctgagggg ggcgtctcgc gattactata acaagcgcct tcctgaccaag
accaagggga
421 aacggttccac ctcaggtttgc cattttcaaca aactgttcgttgt gtcattacccagaagttttg
NM_003120 (PU.1) SEQ ID NO:43

1  aaaaatcagga acttgggtgtg gctctgtcaat gtagggcgggaga ggggtcacc cagggctctct ctgtgtgcttc
tgtacctct ggccaggtggc ctggagagag 61  gttactctcaag ggcccaggtcct gacgccgcta tgggccaag
cccctgcacgg
tgacctctggc
121  gcacacacatc cgcaggggtgt ctgcattgcc cccacacccag gcaggggtgc
tgaccgactc
181  ggcagccccgg cggatgttac agcgcggcca aatggagaggg ttctctctcg
tccccctccc 241  atcagagagac tcgggctgt ctgacacagg tctataacac cggcaaaacgc
cagagatatta 301  cccctatctc agcaggtgattag ggagacccac tagcggaccat tactgggact
tcccccccca 361  caagcgacac agcgagttcg agacgtcgcag ccaacaccaac ttcacggagc
tcgcaggtctt 421  gcagccccgg cagcgacagc agcctctaccc cccakatggag ctggacgcaga
tgcacgctcct 481  cgataccccc atgggtcgac cccatcccaag tcttgcccaac caggtctcct
acgctgcccg 541  gatgtgcctc cagttccac ccctgtcccc cgcacagcgc agctcagatg
aggaggagggg 601  cagacgcggag ccaccccaac cggccctcgt tgcgcggcag gcgcagggcc
tgcggccgcgg 661  gcctgggtgcct gcgctctgggg gcagcggcag cagagaacag atccgccccgt
cacagttcct 721  gttggagaca cctccgccagc gcacattgaa ggcagcactc tgggtgggtgg
cacaaggacaa 781  ggcaccccctc cgtttctcgt ccacgcacac gcaggccggt cggcgcaggtgc
cacaactcagg 841  caagggcaac cgcacaaagag tgacacaacc gaagacgcgg cgcgcacgc
gcgcacagct 901  caagacggcc gaggtcaaga aggtgaagaag gaaacgccacc taccagttcag
ggcgcaggtg
Figure 9 (cont)

961 gctgggcccc gccgggcttg ggcgaagggcg ccaacccgcc cactgagccc gcagcccccgc
go1  cggccccccc ccagccccct cgctggccat agcattaagc cctcgccccg cccggacaca
gtcccccgc
go1  gggaggagcg ccgccccgggc cagaggccagg actgtgcccc gccggtccccc
gtcacccgcc
go1  ctctccccct ccctcaggccc cctccacccc ccgcttcggc ttcctccagg acctccccc
go1  ggtctcggac ggcaggtgqcc ggtcagaccc accggcaacc ttggcacagaa
cgaccgggg
go1  ttcgcctttg ggaattccta gtcgctatgt aatacagact tccccctctc cccccccccc
go1  ccattaacct cccccaaaa aacaagtaaa gttattctcct aatcc

NM_001973 (ELK4, transcript variant a, mRNA) SEQ ID NO:44

1 ttctctgtaa aacaaacacc taatatttttt cttggaggtt ttgttcaagct gtcttaatattt
go1  61 atgactttacc atcccctctgt ggtcataaatc gtcaagtctctcggt ttataagttc
caatctgatag
go1  121 ctagttctctt aaagatattta cttaatgaga acctctaaagc tagaaactct
tgctaggtt
go1  181 ttcatgacecc ttatatttttt taatcattac aacaaactca agatgggttt
cctcctaca
go1  241 tataaatgat gactttttttta gagoagttaaa ggttgcataa aattggttgg
ttaggaggg
go1  301 gtagagacac caaggtgatttt cccttcgcttgc cctccccctct caggggaacgc
tttccccag
go1  361 acctcagccg ttccattaagg ccctccccag gctagaagaac cacagtgcata
tttaggtttt
go1  421 tatcaaggtc ttgtaggtgc ttcggatctg gcaccctcgtt gagaaagctg
tgacgccccc
go1  481 acacactaa acgcagagttt aaggagacac accgacagac gaggagacag
cctcctccca
go1  541 aatctcacc aacacagaaa ttggtcccccgc caggtgcttc tggcagccgc
acccctttcc
go1  601 tccaaaccgg agagaattggg ccggccttggg aggagtccca ggctccccggcc
cagggcaacg
go1  661 tttgggttcctt gggcctttcgc ggacggggcg ccctagggcc tgggaagaac
cgcaacggac
go1  721 ggcaagggaga gcggagcgac caagcagcag ggccgcggacat ctaaactcccc
cggccacacc
go1  781 aacaaccgggt ttggctggcag acggggctcca acgacacccc tagggctcga
gcccesccccc
go1  841 cgggaagtgcag tggtagcgac taagggcttct gggaggacc cagggcagtt
cgggggtttt
go1  901 gggagtgcgg gcggggcggcg agggcctggc aggccccggcc gctgcaaggg
acggccccaga
go1  961 cggcgcggcc cgagcgttag gggtcccaag accggcgcgc cggcgtcggc
gtgccggggg
go1  1021 ggccagggggc ggccggccagg agcccgccg ggcggagagt gcggggggcgt
cggcgcaccc
go1  1081 cggggctcggg ccggccttcctt acggctcggc cgggggggtgc gcagggggtgc

1 1
Figure 9 (cont)

1141 cggcgcgcagtc ttcgaatttc cagcggtgagg aggggctgga gggcgagagag
ggcgcctcgag
1201 tgtcagggag agaccggaggg ggaacccggcc gcgcggccgc gcgcgggcgtc gcgcattgtgct
atggagaggtg
1261 cttaccctct gttggcgagtct cttctcagcct cttcgcagcct gcctcagac
aagcactcctga
1321 tctgtgatgtcg ccttaatgat gggcagttta aagctttttc ggcgaagagag
gtgcgtgcgcag
1381 tctggggtgt ctcgcaagacg aagcctcaca tgaattatatgca caaactcagc
cgaccctcctca
1441 gataactatta tgtaaagaat atctctcaca aaaggaatttg tcagaattttgt
1501 tttgctctctta tccagagagt tttgacatgg atccataatgc agttgggagag
attgagggtg
1561 actgatcagag ttttactcata atcgatctac gcagcttgcgc caaagatgtg
'gagaattgagag
1621 ggaagagataca accactcagag ccattgcagc aagctcctcag cgcgaatgac
tacatacactct
1681 ctgggtttctta tttctctcatt aactctctcct ctttcagact cttcagcatct
aagcctttcctca
1741 aattgataaa gactgataaat ccagccgagag aactgppcagag gaaaaatct
cccctcagagcgc
1801 ccacaccatcctgctcataa tttgtgcacga aacccctccca aagaaccagg
1861 ttgctggcaac cattttcata gtcctttatt ccctcctcata cttcagaaag
acctacctcaag
1921 ctgggtgagag attggttttcc caaacaactgca cttctccctgga aagccacaacc
ctgccctctctta
1981 acgtaatgac tgctttttgcc accaaccacc ccatttctgca cattcccccct
1941 ctggggagagc acctctccac caactggttt ctcccccaca cattgcacaca
gccataatttt
2001 cagtggcttt ctcagccatg gaacctctcag aagaattttgac actggagcct
aaagaccaggac
2061 atcagtcatt gctagaaaaag gacaagggata ataatttcata agaatcagag
aaaccccaagag
2221 ggtaggaact gcgcccaac cttgtgtcaaa cggagcgtgaa cccagcccaacc
ctgggaatcac
2281 tgagccccccttctccctcata gttctctattt caacccgactt cttccccagag
acccctctcatc
2341 tactgcattgg cagccctccttg cttctccagtag tccacctttctt ggactctctc
agtcctggttgc
2401 ctccccctcag ccagcccaag cgtgaaggctg ctaacacacct ttctccagttt
cctctctctcag
2461 tgaacagctca tggcagcatcc atctcttgctg ggctggatgtg acctcctcacc
ctggcgcctaacc
2521 cttcccccaaa cctacaagag acataacata tcgaaccttggt gcaggagaga
acggacacag
2581 gaagaaaacgc atatatctcctag tgaagttgagc aatggagatg
ttcacattgcc
2641 tggataatgac ctattgtgtgatt tttgcccatt cccatggagaa acatatccttt
1tagattctct
2701 tttgaaaattg ccactaagttg gactatatgt ataattatgc cttatttgcgtcttaactc
NM_021795 (ELK4, transcript variant b, mRNA) SEQ ID NO:45

1  ttctttgta aacaaacacc taatattttt cttggaggttt tttgtcagct gtoctaatatt
 61 atgactttac attctttctg tgtctaaactg gctcaagttag cctcttgtag
caagtgtgac
 121 ctgattctctt aagaattttta cttatgaga accctaaagc tagaattactct
tgctagttgt
 181 ttcatggacc ttatttttct taatcattac acaactcaat agatgggttt
ctctccacct
 241 tataaatgtat gactggtttta gagagtttaa gttgtttaaa aattgggtgag
ttagtggagg
 301 tgtagagccac gataggtattt ctgtgctcctg cctccatcgt cagggcaagc
tttcccacg
 361 actccagcgc ttccattttg cagttcccag gcttagaaagc cacagtgcta
atttagtttt
 421 tatcaagcgt ttgtaggttc cttggatctg gcaactccgt tagaagctg
tgacggcacc
 481 aacctttaac agcgagttgt aagaggtagc aagggaaacag gagggaggag
actctcccca
 541 aagcttagcca ccaacagaga tttgtcccccag cagttgcctc tcgagcgcgc
acctctcccc
 601 tccaacccag gagaagtttg cgcccttttg agggattccta gttccggggc
cagggccgag
 661 cttggtctct ggcgggtttc ggaggggagg ctcagggacc tggagcaaac
cgcaccggaac
 721 ggcgcgcgga ggcgcgcgca gactccagga ggcccggcag ctacatccccc
cgcaccggac
 781 aacccgggtt ttgctggcag acggggtcct cgaacccccct tagggctgcc
geccctcccc
 841 cggaggagcat gttgtagggag tggcggtcgt tgtggagggacc aagggcagt
cggggggtttt
 901 gagatgaggg gccggcgcggc aggacactgg gcggccggc gcgtgcaagga
agcgggggaga
 961 cgccgggccc cgggggtgtag cggcggccaaag acgccgggcc cgccggtgcc
gtgccgggggc
1021 gggggagggg gcgggccccg agcggggcgg cggggagatc cgccgggggc
cggccaccgcg
1081 cgccggtcgg ctggcccccc ggcgcctctct acggcctgcc gcgggggggtc
gcgccgctcg
NM_005240 (METS (ETV3)) SEQ ID NO: 46

1  ggggggggtg gatgagagag aggccggagac ggcccggagg agacccggacc gaagacgac
   61  cctgcctgqa agacccggqg ggtgaaatag aagccggcgt gtagcatacg
gaaaaagcc
  121  gaaggaggtg gagggtatca gtttcttgac tgggcttaca aacacagatc
  atccccggcg
  181  tccggccaga tccagctgtg gcacttcata ctggagctgc tgcacagagga
  agagttccgc
  241  ctggctacag cctggccagca gggagagtac ggggaatccg tcaacagga
  tccagatgag
  301  gtggcccggcct cttgggccggc caggaatgc aacacccaga tgaattatga
   caagctgagc
  361  cgggcccctca gataatatta caacaagagg actcttcata aacaaagg
   gaaaagatt
  421  acctataaat ttaacctccag caagctgttggt atgcoccaact accoaattat
caacattcgg
  481  tcaagtggta agatacaacac tcttttggtta gggattaat tttgaactga
   aagaatttt
  541  taaaaatcaca satacagac atggcatggt taggaagatt taggaacac
   taaaatagtt
  601  tgatccttttg gattgcctca atgtctctac tcaagtcata tcaacttttaa
   gagnagtta
  661  agggctattca gatatactag agatatttctg gtttatatttg gttgtctgcc
tttctgcat
  721  atgtttttatga atctcctag aatatttactg cttctcttg gttggaatgat
gttttcatag
  781  ttgtgtgatga tcttttcggt caggaactcag tttaaacacc cagccagctg
   gttctttcct
  841  agatgggaac ctttttctac aacacctcag gattttcctg gaaactacca
   agctctccct
  901  tatcaagtga atatcatcaca aaccacagca tctctgctga gagaaggggg
   aggttcacat
  961  gtttcgaaggg aaaaagacgtg ttttggtatt gcacacagca atcccaacag
   aaaaagatcc
  1021  tgggtctact tgaccccttc tccgtttaag tgcagtaggg cttccccctct
tgacctttctc
  1081  ggttatagct tccctactaca gttccccaca tttactctttg atggtgaaag
   cagtccccca
  1141  aagacttttg tgtgtgtgtg gttttttgtt tgtgattttt ttccttatgc
   aatcatcact
Figure 9 (cont)

1201 cctgcccaag aaaaatacagt agttcctctt atctgacgcag tatatgttct
aagaccccta
1261 gttagattcgc aaaccacaga taagtacaaaa ctccattct atatatgtg
1321 ctctactcccc atctatagtg atctgtgata acgtaaaatt tataaatag
1381 agattaatga caataataaa atagaaaaat tataaaaaa aaaaaaaaaa
1441 aaa

S72620 (EWS/Fli1) SEQ ID NO:47

1 cccacctgtt acccaccccc acctgcatcc tacagcccaag ctccaagctca atatagccaa
1201 cagagcagca gtacgacgga gcagaatccg tacagatccg tgggccc gagc
cacagtgcgc
1321 ctgagcaaccccttgagacc gcagatccag atgtgctaatctctgtgaa
gtctcttct
1441 gacagcgccca acgcacagctg tatacctgag gaggagacca acggggagct

BC029743 (ESE2 (ELF5)) SEQ ID NO:48

1 cacaaggtcta cagtgctttt tattttcact gcacgctgtt gcgtggagcgc cctgcctttct
1201 cttgcctctg aagctctctct ttggacccct gcacccgctg ccttcacggt
1321 tcggtgacac acacacccctt cctgcctaatt gcatactttt gcgcctcctc
gtctgtgcac
1441 acgtgactgct tcagcaatga agatctactc cctgccttttga agcatcaagac
1561 cccatgtgctc
1681 tcatactgga catcagtcra cccctgaatag tggactaagc gcacagltgtg
1801 agttggtgctc
1921 gggctgcttct
2041 cagttctctct gcagacagta caggtggac ccacattgcct ttcctctttt
2161 cgaacccatac
2281 atcagtgcccg tgcagctgtg cagcatgaca caggagagct tgcctcgacgc
2401 agtctgcttct
2521 tgcgtgagcg agtctgacttt catctctcag aacacccgca cacaaggtta
cctcttttttt
2741 aatgacgcttg aagaaagcag ggccaccctc aagacatagt ctcattccaa
2861 ctgcttgaaaaa
3081 acaagttgcca tcaaaattca agactgtcag tgcataatga gaacagcctt
ccaagtctttc
3301 cattctattgg aatatttgacg agacagtcttt catcttctctg aagaaaaacctg
tggcattctg
3521 gaatgggaaag ataggaagca aggaattttt cgggtggtta aatcgggaagc
cctggcaaaag
3741 atgtggggag aagagggagga aagagcagag atgcagcatatg aaaggttgag
cagagcctcag
3961 agataactct ataaaacagg aatgtttgag cgggttggacc gaaggttactg
tcacaattttt
4181 ggaaaaatgt gcaacgggtgtc gcagagacag aagctgtatag tgcctccagg
catcaagctc
4401 attttatgta ttctgtcttt ttaaaaaaat cagattgcac tagacattcag
aaggtcttca
Figure 9 (cont)

181 tggacagaca gctactcaca actsagatgtt tccagtgggt ttttggaggg ctagatggtt
gcacgcctct 241 gaaattctc ctcagtaact gaccaagtac caaggtggg agttgatcga
gcaccctcct 301 gacaccaacc agctggtagc caatgttatc cctttcacaag agttcagcat
cacggtggag 361 cactctgcag cgtactagtt gacaggattc acccgggccc caggagcggc
ggggagacct 421 ccttaaagca ctgctgacagca tctgaagtgg aacgagcagg gcaagaggtga
cctgtctcag 481 tcaacacaca atgtcattgt caaatgtgaa caaactgagc ctctcatcag
gaacacctgg 541 aagacacagag aactatattata tgaacaccaac tattgtgac gatagatatttt
gttgacagc 601 aaaaaatctct ccggggtctca gatctcctgc acaacaccaac gtcacctccc
tgttgacagag 661 tcaacgtgata gtaaaagag agcagacaccc cctgccaagt gccacacacaa
aagcacaac 721 cagagagggg ctcacattct gagatttcac ctcgagcactc cttggaacct
tagcaacacaa 781 cagagattaa taataatgggg aacgagctct ggggaggtct ctaggttttt
gaaatcagag 841 gtcgtggtct acgtattggg taaaaagaag aacacagca gcatgacota
tgaaagctc 901 agggcagcta ttagatatca tcaaaaaaga gaattctgg agcgtgtgga
tggacagao 961 cttgttattata aattttgagaa gatgccccga ggatgagag ggaaatgaaaa
cTGAGTGC 1021 caatacttgg gacacacacc aaaaacacac ccaataatac agaaacaaag
aactctcggg 1081 cgtaaatatt tcaagacta ctttctcttg atatattag accatgaggg
gacaagaaa 1141 ctacactcaca cgggaagag aacactaca gtcgattaata aaaatatttt
ttactctcag 1201 aagtatgtcc tatatgagga aaaaaagtac acagtttttctgtgaaatag
tagctgtatg 1261 tggcttctgat tttttttcac ctcttatcttg aatctttttt caactgcaaga
gtaacaggag 1321 ttgtagcttt gttgcttcttg ctaagagaa gaaaacaaaa atcagagggc
attaatattt 1381 ttgtagatgta cagatttag aaaaaggtga tgcacagaaa ctaacataaagc
atccatatgg 1441 ctcctgcaag ggaggtgac attctgctgt agttactac cagggctctca
gatgtgagg 1501 acaaaagtggag tggatgcggc gaagttaaa ctcgagctta ggtcctagttt
agtagagaat 1561 ggggactttcc aaaaacccag gtttgctata atctctgcat aaccacatga
cctggagtgc 1621 ttaaaatcagc aagagaata atggtggggt ctttatactc attcaggaat
gggggtcag 1681 atggccagggc tgtctttttt tctcccccttt ggtggttggtg taatttacctt
taaattctct 1741 gtctgtgtccca ttcataatc ttaagagaga accacagttg gtcttcttttt
gctcagagttg
Figure 9 (cont)

1301 cttaaaata agttggaaaa aggagacggt ggtgtggaaa tggctgaaga
gttgtcttt
1461 gtatocctat agtccaaagt ttctcaatct gcacaattga catttttggc
cggaggtggtc
tctagtctc
ttctgtcctct
ttagatggtc
taaccacttg acaccagaga ccaccagact gttgaacgc aaaaagtctc
tagacatcac
ttaggtcttc
tacatatgtaa tgggtgttaa ttccgaaag aaaaaagata tgaaggtcttt
ttagctcttt
taggttaggga
agggacttgg aagatagtag aggggctttct tttcagaacc ccagatgaga
tagaatagtc
tagaggtgtc
tggataaaa gctagtagat gtagcaggaa ctacaataga agacatttgc
tctgaatta
tataaaatgg tgaaggtatt ccaacagag taaattgatg
tccagggaaa
ttcatctgt
tatatgtcctag atctagcccag gtggacgggt tatggtcttag atttgtactc
agaggaattt
ttttttcttt ttttttcttt ttagggaaa
tccagaaacgt
ttagaatgg
ttttggtttct ccttacatct cccttccttt atctaccttt cttcagcaaat
aatagcgaat
ctcataaaa tattgttgtct ttagcggact acatcggc aaaaatga cagaatttggctt ccacacgt
ttctgccaaat cccacatgatat ccctttcct cggattaagtt tgggtacgt
tgttgataata
ttcagggctctgtta cttggaggttta aaaaaatgactgag ctcacaataac gagaagagtctg
tatgttctgcttttctcgttttcttgacgtgcttggcaggtcag cagctgtttttcttctggttttttctcgtgcagttttcctcctcttcttgagctttttgctcctgggttta
ttcagggctctgttaa aaaaatggtcctt tttgggtttt ctcttcttctc ctcgacgttttctgacgttctttcctgtggttttttctcgtgcagttttcctcctcttcttgagctttttgctcctgggtttta
ttcagggctctgttaa aaaaatggattg cgcctttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Figure 9 (cont)

3421 ctagatccttc gctgacacttc accctttcacc taacctgttaa gtttctcactctttccaaat
3481 tggccttggcactctttccaat tcaagagc ttcctttgttaa
3541 aagtgaaga gtagaacaact caacactattac acagataatt atgtgattgt
3601 gagagtaag agtttctt

FO 3S E. g. tctgagttg (Casag & C Co. 6 agtocticcaa. tga.cgg to 2. c cago agg ca...

AP071538 (PDEF) SEQ ID NO: 50

1 gctgacttccccctcgcac agtctgctcac ttctgcttgt ccacactgcccaccagaccc61
 61 agtcctccaa ggcctctggcc agtcctctgc aagcccccctea ggttggtggtgttgctcggctagttggct
 121 ccacgaccgcc gcccctgggct ggggtttaggg gactccctac aggcaagcag cccttgaggacc
 181 tcagagggcc acccccttggag ggtggggcag cccccccttgag ccaacctgagcagtgctctctctcgtgctcgtcgtgcccctgagcc
 241 gcccacaggcc ctgcttggccc ctggttccgg tggccccccca gatgctggtgc
tgagacacgc301 cagtggccttc agctggccccc acctctttcccc ggccccctgaa gttggcaactg
cagcagacag361 cccccctgggaccagggcag ctagacacac acgcgggcagc ccacacagca
gcggccagggg421 cagcgccacg cccggcttcga gcagctattc cccacgccccctccttgctgccc
cccccgacacc481 ggtgtgtcggg agaagcggttg gagaaggggcca gtcgggtctccg
agagagggga541 ctggagtccc agtcctaccccg ccacgccccaga gcagggcttg tccgcctctctactcttcctctgtggcttagcagagccgggggccttggc
taggaggggtcgaccacagc501 ctttgacatcctgacttctcg aggacagcag ctgggaggac ccacacagcag acgcggggccttggctcgcagtgctgttggg
666 ggccccgacac721 ggccccgacac gctcttgctgt cagcttggtgct gcagggggtgt gcattggagg gacactctgc
tggagcaggt781 gcagctctgtggtggtggggtg aatggtctccaa gcagacatcag aggcgttcgca
agcgttcgca841 cacacccgtca gatcctcatggt actggagcccg cagcaatgtg cagaaggtgc
ccacccgacgtc901 agacagacag taacggctgc cccccatggg ccagcccctcc caggagctgg
cggcagggag961 ggtgtgtcggc agttgctggg agcaggtttccgcacgccttgtcg ccccctgtcgg
gggtgtgtgt1021 ggcggcaccac ccggtctgcttg ggaagtcacgc gcgccttgatg aaaagcggga
cctcacttcgg1081 ggcggatccag taactgtgtgtg cggaccactgta gggagcctgg accacagcg
aggtggacagt1141 atcatgtctcc gggcagccca tcacccctgtg gcagttcctc aaggaggttcgtactcagc
caacagcatctca1201 gcagcatatgg gcggctttcatatctgccagagagtaggggctagttgttttacttctcaaatgga
Figure 9 (cont)

NM_005230 (NET (ELK3; SAP2) SEQ ID NO:51

1 ggggccgaaaa gcctggttttc acagactgca cacgccttgg ggaataatgc agtaaaggaaa
gaggggaaaa
2 aaaaagaggg cggagagaaaa gaaaaaaggg ggggaataac agaatctcat
tacaagacgac
gcttgctgcc
3 cgggcggctc tggagtctgctt gtggagattg ctgtgccatg gaggcctgcc
4 aattcsctgggg ggtggctttgg gtagctgttg tggagagagct
tgtggctgcc
5 cggccggttttt cggaaagatgg cgggtgtccct cgggttttttt cgggtttttt
tgtggctgcc
6 ggggtggtggct cgggctgttt ggctttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Figure 9 (cont)

901 gcggccggt tccgcttctt ggcctctgcc gtcctggcec agatctctctc
ttabatttg
961 ccacaagctg ccagatttta atccggctca cccctctctt ctcggctc cc
gtctcgtc cc
1021 ccacaactca cccttcccttc tgaacacaga agcctctttc tcgggccgc gc
gtcccatgac
t1081 tccgatctcc cggacccccct gaacctgtca tcggctctca agacacagt tc
ccactctctt
t1141 ccacaacaggg cccaaaaacc caaaggtttg qaatactcag ccgccccggcgt
gtgctctctc
t1201 gcggccgaca tcggctctct cgcctctcaac agcggcagcc cccctcteggg
cacctctacc
t1261 ccagctctct ccagcgcaca gacacaaat ggcattgtcc tcgactcgcag
tccactgctc
t1321 tcagcataac attactggag cagccttagt ccagttggc tcgctgaggtc tc
tgcaggttt
t1381 ccagggcaca gcacgctttg ccaagttcccc acaactgctta atggccacat
gccaggcga
t1441 atccgcaagtc tggagacagc gtcattttcca gtaactgcttt cttacaactc
tcgaattttc
t1501 tcgagacgct cggcccaaat taaggactca ttaactgtag taaaacaaat
1561 gctggctcac ccctttcggt agagaagcat tggtaacctc tttttaaittt
gttttgacat
t1621 ttctcataca cttgatagtc atagttatgt tagcattatta aaaaacgttttt
tttgatattt
1681 ccagatata taagatcttg tttgcatca agtgaatttt aatgttttgt
ttttttatc
t1741 ctcttgagtc ttaagtggtg aaacagtttg acagtgaaga acttttctta
atggtttttc
1801 gtataactta taaggatgg aagctttttt ctctttagtt tctgagttgc
ttaaacggt
t1861 gttttatag actataacca gtttgctctt cttttgcatt tataattgaaat
gaatatttt
1921 tatatttttt attattttaa ggaatatttt gaaagatgaa aatattgatttc
aaacagctct
t1981 ctagtagaat ttccattttt ttccacagtg ggcataatga aagcatatat
cacgattttg
2101 gtctaatattt gcggctttttc taaaactgtt gagaactgttt ttaaactgaa
2041 ttacttttta attgataag aatgtgcttt gacccctatt tgaactggaa
tttcagtaat
2101 gtccaggttaa tgttctttata ataactag ccataattag acaataaaaa
tcggggggg
2161 aaaaaaaaaa aaaaaaaaaa

NM_006874 (NERF; ELF2) transcript variant 2 SEQ ID NO:52

1 gttgccagct gcggccgccc cccacaagcac acgcgcgcgc gcggccgccc gcggccccctg gc
61 cccgcggccc cctgcgccttg cegtcaggg gcgggggttcc tcagcgcggc gc
gcggccggcg
121 gcggccgggcct ctgctcctgc tcggagggagc cgcaaggactc gcggggtcgg
gtgggtgtgc
181 agtgcggtaag tgtgcgtcgg tcggcagcgg cgggtgtctcc gcggcggttt
cgcgccttctc
Figure 9 (cont)

241 ccctgccgcc ggtgtctcaag gtgtgaagtca atgtgaagca gcagctccag ccctgggata
301 aacatgggca ctgtctctgca tggagggcacc acaagaacagc tggagtcgtct
ccctgccc ggtgaagcac cagttcacaag cagtaatgca cactgtcag ataagacaat
tgagatgtcct
421 gaagccctgca ttcatatgga atctcttacc tggtagaggag attcaagaag
tctctgtggaa 451 gtgtttgttt cctctctgtgt atcaactcaca gaattcatcc atgcgtgctat
gagccagat 541 gtcattagc aacgtgtatg ggaggtgtca actgaagagtt ctgaaccctat
ggatatctct
601 cctattcaca cactcaccaga tagccctgaa ccaatgaaaaa gaaaaaagat
tggcgtaaaa 661 caagagaccg agcaatcaac aatttccaat gggtctcttg agttaggtat
aaagaagaa 721 caagagacag gaaagagga cacaacactt tttgtgagag ttcctttaga
tctcttccaa 781 gataaaaaata cttgctccag gtataaaaaa tggaacctaga gagaagagg
catatcacaq 841 ctggttgatt ccaaaggcgtg ctctagccttt tgggggaaagc ataagaaaaa
accagacagt 901 aactatgaaa ccaggggacg agcttttgaga tacctactacc aagggggaat
tctggcacaag 961 gtgggaagagc aagggctttg atacaatgttc aaggtatgac cgaaaaaacat
agttctcata 1021 gatgatgaca aagatggacat ctgtagaag aagttgacag gaactactga
tgaaaaaata 1081 ttgagaacag tgctacgtgc tgcagaaagtt ctctctgaag cagcatcctc
tgtgcgcaag 1141 ggaaaaaat cattccctat aacatgtcctc agagcagaga aggggtgtagc
tagaggttggt 1201 aatctacctt cccctgggca cagatctcttc tccaggttctc ctactaccac
tgacatgtgc 1261 tccacgccag cagctccaaag gacagttcctg tgtgcaatgc agttacctgt
tgattgtgaca 1321 tccatgggtc agaaaaattt aacgtgtggca gttcagtcag ttaattgcag
tgcaccatta 1381 atacaccgca ctagtccaca acacgcgaacc tctcacaaggg tagtcatcctc
gacatcctct 1441 acgtgtgatgc cagcttctac tggaaatgga gacaaataca ccatgcagcc
tgccaataact 1501 attacccata cagctacaca gttgcaacag tgttaactgca agacaagagtc
aatctgcagt 1561 ggagcagggga gcattaacagt tgttggaacc ccatgggttg tgtgaagcact
taccctgctt 1621 tcaatagggc atggtagacc tgtgaatgaga ctatcaatgc ctactcagca
gacatcctgc 1681 cagctttctcctcagttat cagttcagtc ataaaggggc cagaggtaaa
atcggagaca 1741 gtggcagaaaag cagcaagagag tcgagttgaa acttgcacgc tagtagaaga
aaaaccagca 1801 gatgggaata agacagtccag ccaacgttagg gttgctagtg cgccttcagc
tattgcccctt
NM_ 201999 (NERF; ELF2) transcript variant 1 SEQ ID NO:53

```
1  aaaaatgtga aaggatgtta gactactttaa catacaaatc gctttctggt taactctcttt
 61  tgaagaacct gatttttgg aatatccactc tattgacctttc ttaaaaactatg
 121 tatacactct ggcacacact gctggtggtgta atataacttct cctgcccttt
181  gttgaaggctt ggttcgtcctag aatagcagcatg ctggactccctc tcaaatatat
```

Figure 9 (cont)
Figure 9 (cont)

241 ttcttggagct ttccagcaat ggagtagaaa atcaagagga aagtgaagag
246 gccactgat
301 atccacgctg gatttgaggg ccagttcaaa gtcgccagatg aagcaggggc
tatgcagccc
361 aggtttcttgt ttatgatgat gagaactata ttagtgcaaga gtatggcagaa
gaaccaagaga
421 ttgagacgga gaatgtggaa acaagttgaa catcagttca cagcagtaat
gcacactgta
481 cagataaagac aatggaagct gcagaagccc tgtgctcatat ggaatctcct
acctgcttgga
541 ggattcaag aagctctgaa ttcatccact gtgcattgag gccagatgtc
attacaaaaa
601 ctgtaagttga ggtgtcaact gagaagtctg aaccatgga taacctctctt
atccaaacat
661 caccagatag ccatgaacca atgaaaaaga aaagaattgg ccgtaaaca
agacaccagc
721 aatcaccaat tttcaatggt ttctctgaag ttaggtataaa gagaagaacc
agagaagggaa
781 aagaaaccac aacctatttt tgggagtttc ttttagatct acctcaagat
aaaaatacct
841 gtcgccaggtg tattaaatgg aacctagagag aasaggtcat aatcaagctg
gttgatccaa
901 aggtgtctct taagotttgg ggaaacgata aqaacaaacc agacatgaa
tatgaacccaa
961 tggacacgac ttggagatac tactacaaaa ggggaattct tgcaaaagtt
gaagacagaga
1021 gctattgtta tcagttcaag gatatgccga aaaaattagt ggtcatagat
gatgacaaaaa
1081 gtgaacacttg taatgaaagat ttagcagggaa ctactgatga aaaaatcatt
aagacagtgt
1141 cactgtctgg agaagattct ctagaaagac catcctctgt tcgcagtgga
aaaaaatcatt
1201 cctctataaa cttgccttag acaagaaagg gttcgaagtt aagttcgaat
atcactttccc
1261 ctgggcacgga tgctctcatcc aggtctctta ctaaccactgc atcgttgtaa
gcacaacagcg
1321 cttcaagagc agttcgtggt gcacacgagg taacctgtgtg aatgacataca
tgggtcaga
1381 aatatttcaac tggtgccqatt cagtcagttta atgcaggttc accattaata
accagacacta
1441 gttcacaacc agccacctct ccataaggtag tcattcagac aatccctact
gtctcagccag
1501 cttctactga aaatgagggc aaaaatcaca tcagagctgg ccagattatt
accatcccag
1561 ctacacagct tgacacagtgt caacctgcaga aaagttcaaa tctgaectgga
tcaggaacca
1621 taaacattgt tggaacccca ttggctgtga gagcacrat cctggtttta
atagccccatg
1681 ttagcactgt gcagatcact taatgccta ctcagcagag atctggcagc
tacctctctcc
1741 gaagttatcag tgcagtcata aaggggccag aagittaaatc ggaagcagtg
gcacaagaaagc
1801 aagaaacatga tgtaagaaact ttgcaagctag tagaagaaaa accagcagat
ggaataaga
Figure 9 (cont)

131 cgctctccccg gctgcgggge gcggggggct gcggggtgcc cttggtctgtg
gggcgggccc
241 tttgagacttt tattgcagtg gacqgataag aggggcgggg gggggtctct
gggggccggg
301 gcggcagcgc ttttaattaa acggaatttg cgccccccggg cccgcggggg
ggggaggggt
361 tccaaaggggc cctctagcgt gaacgcgtttt tctagagaaa cccgcaacc
ggggctgccc
411 cgggggtcc cctcccccggc gggcggcggc ggagccccggt ccgtctcttc
tctcctgcga
481 ccgccccgcct ctcgcccccg ggatcgggcgc acacgcgagc ccagccctgcc
tccgcccccct
541 tcccccgcac ccccccacttcc cacggcccaag tccccggggg cgatgagaca
ggcggggcc
601 tccccagccc tgcgtgactaa ccctgcttttt ccagatcccg tccgagaggg
tcccttcaagq
661 gacgggaaga aaccgagcgtg gggggcgggt agcggggcggg tccagaaagg
cagggagacag
721 atcagctggt ggcaagttctt gctgaggactg cttggtgcacc gcggagacgc
ccgagctcgc
781 gcgtggtggag cggttgctacgg cggttcaag ccgagcggagt
gggcgccggg
841 tggggcagag gcgaagagcga gcggcacaatg aactagcaac aactgagccg
cgctctgccc
901 tactactagc acagaacacat catcagccag cttgcagggca acgagctacgc
cacgcttcgc
961 gacctccagg gcctggcgaca gcgtggcgcag gcggggcccg cgcaagctca
tgcggccgcc
1021 gcagctctctg ccgggcgcgc gcggcggcag gcggcccgcc gcgtgagacgc
tccgcggggc
1081 ctggcgcgccc tgcctctttcc ccggctcttcc aaactcagcc tcatggccgc
tctggggcc
1141 gtgcgggcgc gcgcgtcacc ctactgtgcgg gcgcggggcc cgccgcccac
gctgctgcgc
1201 gcgcggccgc cgctgccacc cagtcgaggg ttggaaaggg cgcgggggccc
cgggggggccc
1261 ggtggccggag cctcgcacttt ggggggccat ccacactaga ggggcttggtc
ggggtcctgc
1321 gcctgcggcc gcacgcagtag gcttgtgccc gcacectagc gcacgggagc
gagggggccg
1381 gacgcctcgg cacgcgtcct ctaaatccaga gtttaactcc cctgcgcgcc
ctggacaggg
1441 gacgcgagcc cagctccctct aaactcgtggc tggtaactga tttgctcttg
tggccacccc
1501 cagttcccttg aggaggccga tgtgctgcctt ctttcaacttt tttttctcta
ggtctccagg
1561 tccggggaggg gatttttgga ccttcctcttg tcccccacca ctcagctgca
ttcgggccctg
1621 tcctcctgag cccccatcga atatactaca ttenttcaagga gtgcacacac
tttttccact
1681 tgtgcctttc ccacaggaac tggcctccccc tccagcagcgtg gaggctcctc
dcgtctcttc
1741 tccttgagac tctgacaggtt tgtggtaaag ccaacgtctt ccgtgacaca
cgcccccttt
Figure 9 (cont)

1801 cctctgttc ccacacttc aggagaaact cccgtagtgt ttctgacct ttctgcctca
1861 ttaaagctct ccagctctca aaaaaaaaa aaaaaaaaa a
Figure 10

NM_005656 mRNA (SEQ ID NO: 54)

1 cggagctgaa gccggagggc gggccgaggg cggagccgaga gggccgggga ggcgcgcctg
gtcctactg aacatctcag atacattcgc itactgtagg
cggtgtgataa
121 cagcaagctg gctttgaaact caggctcacc accagctatt gcacccattc actagtcaca
atgaaacaca
181 tgtgatacc acccgaaacc cctataccgc acagcctacc tgtgtcccca
cgtctataca
241 gggtctacgg gcctcgtact accggctcacc cggaccccaag taacgcctcga
gggtctcgtac
301 gcaggtctcc acccctgatcg tctgtacaagca gcctaatctc cctacccgga
cagtgtgctaca
tgctgggagc
361 tctcaagagct cagaaagcac tgcgtcatac cc Tgctacagct gcggaccttc
cgtctgggagc
421 tgcgcgtggcc gctggtcctac tctggaagtt catggggcagc aaggtgtcaca
acctctggagat
481 agagtgcgac ccctcaggttag ctgcatacaaa cccctcaaac ttgggtgtag
acccacacaca
541 ctgcctccgg gggagaagac agaatcgggt tgttcgctct caacggacca
acccagactt
601 tcaagactgct ctactcagga ggaagtccttg gcacctcgtg tgtgcaagaagc
gtgggaacca
661 gactacaggg cggcggtgacct cagacggaacct ggcctataga aataatcccc
gtggagtacga
721 ggagtagtgtag tagacagacag aacctcaccct tttttagaaa atggaacacca
gtgccggcaac
781 tgcgtgatacat tataaaacag tgcaccacag tcagcgtcttg tctctaaag
caagtgttctc
841 tttagctgttct atagctctgcc gggtaaact ctgcatacaac cgcgcagaca
ggatgtgggg
901 cgttgagacgc gcgctcggg cggcctgggc cttggcaggtc agccctgacag
cccaagactt
961 ecagctgtgc gcaggtcctcc ctcatacccc ccaggtgttac gcagcagccg
cctactgctg
1021 gaaaaaccct ccataacact ctagcgcattg gcagccattt gcggggattt
tgagacatcc
1081 tttcatgttcct ctgagagccqg cattacaagtg agaaagtgct atttctcact
caaatgtga
1141 tctcaagacc cagaaacatcg acatgtcggtg tgaagacgta cagaaacctc
ctgactttcaac
1201 cgcacctagtcccc acacccatgtgt gctgtgaccac ccagggcatc atgcctgcaac
cgaacacacgt
1261 ctgctgtgatt tccgggtgagg gggccacggg gagaaggg cagacctcag
aagtgctgaaaa
1321 cgtggccaggg ttgctccaaatgtttgagctct cagatgaaac ccagatagatg
tctatgacaa
1381 cctgatacaca ccagcctatg caaagagccgg cttctctcag gagagatcctg
atctttgcaac
1441 ggggtacagt ggaggctctcc ttcctactacg caagacatatt atctgggtgcc
tgtatgggga
1501 tcaagcctggg tccctgtgtggt gctgcaagacaca gcctgacaca ggaagtgtagt
Figure 10 (cont)

1561 ggtatccag gacgagatt atcgacaaat gagggcagac ggctaatca catggttcc
gtcctgacg tcttttaca aqaaqaaaact ggggctggttt tgttctcccc
gtgcgatggt
1681 tacacacacg gatgcatcag aggtcacttc attttattta saccagtaac
ttcagttcc
ttgcaacatg
ttgcttattc ctggccatttc gactgtat cac tggctic atgaagatgag
tactgtat cac tggctic atgaagatgag
ggctaatcC& ttgct tcc.ca &a Caggaac C C Cats. gCaig Cig tgagt cottt at gaCttgag gget gcc.ct ... gtggittg: gtottgcaat Ccaggacct atta actgtttgtc. Catgcaaata 3.33 titt. t.tttitt, aaa.gcaaaaa aaaaaaaaaa aaaaaa.
Fig 13

A.

Localized PCA

HN LN METS

TMPPRSS2:ERG Rearrangement Status
- Rearrangement with deletion
- Rearrangement without deletion
- No Rearrangement detected

n=118

n=17

B.

Pathology Stage

Percentage by Category

n=19

n=14

n=26

pT2

pT3a

pT3b

4

5

14

3

15

26

17
Fig 14

Chromosome 21q22.2-22.3

Centromere

Telomere

2nd generation probes

1st generation probes (ref. Tomlin et al, Science 2005)

additional signal on chromosome 2

scale:

ERG

500 kb
ERG Break Apart Assay: ERG rearrangement LN13

No Deletion
10% Cells

Deletion
90% Cells

Figure 15
Fig 16

Oncomine Gene Expression Evaluation of Benign vs PCA Chromosome 21q22.2-22.3
Figure 18
FLI1 overexpression without fusion transcript
Figure 19

Induction of ERG protein by Androgen in TMPRSS2-ERG+ Cells
**Figure 21**  
_Nuclear Interactors for ERG2 (Mass spec and Protein Arrays)_

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Gene Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRP8 pre-mRNA processing factor 8 homolog (yeast)</td>
<td>PRPF8</td>
</tr>
<tr>
<td>T-box, brain, 1</td>
<td>TBR1</td>
</tr>
<tr>
<td>EBNA1 binding protein 2</td>
<td>EBNA1BP2</td>
</tr>
<tr>
<td>V-ets erythroblastosis virus E26 oncogene like (avian)</td>
<td>ERG</td>
</tr>
<tr>
<td>Fibrillarin</td>
<td>FBL</td>
</tr>
<tr>
<td>Exosome component 7</td>
<td>EXOSC7</td>
</tr>
<tr>
<td>Friend leukemia virus integration 1</td>
<td>FLI1</td>
</tr>
<tr>
<td>Nuclear mitotic apparatus protein 1</td>
<td>NUMA1</td>
</tr>
<tr>
<td>Poly(rC) binding protein 1</td>
<td>PCBP1</td>
</tr>
<tr>
<td>Small nuclear ribonucleoprotein 70kDa polypeptide (RNP antigen)</td>
<td>SNRP70</td>
</tr>
<tr>
<td>Lectin, galactoside-binding, soluble, 3 (galectin 3)</td>
<td>LGALS3</td>
</tr>
<tr>
<td>RNA binding motif protein 23</td>
<td>RBM23</td>
</tr>
<tr>
<td>TAF6 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 80kDa</td>
<td>TAF6</td>
</tr>
<tr>
<td>Zinc finger protein 306</td>
<td>ZNF306</td>
</tr>
<tr>
<td>Ras association (RalGDS/AF-6) domain family 5</td>
<td>RASSF5</td>
</tr>
<tr>
<td>Calcium/calmodulin-dependent protein kinase 1</td>
<td>CAMK1</td>
</tr>
</tbody>
</table>
Figure 23

ETV1 (NCBI)

MDQFYDQQVP YMVTNSQRGR NCKNEKPTNVR KRKFINTDAI HDSEELFQDL SQLQETWLE
61 AQVPDNEQF VPDYQAESLA FHGLPLIKKK EPHSPCSEIS SACSQEOPFK FSYGEKCLYN
PHSPCSEIS SACSQ

121 VSAYDQKPQV GMRSNPNNPTP SSTPVSPPLHH ASPNSTHTFPK PDRAFPAAHLPP SQSIPDSSY
TP SSTPVSPPLHH A RAFFPAHLPP SQSIPDSS

181 PMDHRFRRQL SEPCNSFPPL PTPREGRPM YQRQMSEPNI PPPPOGFKQVE YHDPSYEHNT
241 MVGSAASQSF PPPLMIKQEP RDFAYDSEVP SCHSIYMRQE GFLAHPSRTE GCMFEKGRQ
301 FYDDTCVVEPE KFDGDIKQEP GMYREGPTYQ RRGGSLQLWQF LVALLDDPNSH SHFIAWTRGR
361 MEFKLIEPEEE VARRWGIQKN RPAMNYDKLS RSLRYYVEKKS IMQKVAGERY VYKIFCDEPA
421 LFSMAFDPDNQ RPLLKTDMER HINEEDTVPL SHFDESAMYF PEGGCCNPHP YNEGYYV
Figure 24

FLI-1 (NCBI)

1 MDGTIKEALS VVSDDQSLFD SAYGAAAHLP KADMTASGSP DYGDYPHKINP LPPQQEWINQ
   SLFD SAYGAAAHLP HKINP LPPQ INQ
61 PVRVNVKREY DHMNGSRESP VDCSVSKCSK LVGGGESNPM NYNSYMDEKN GPPPNNMTTN
   PVRVNV RESP VDCSVSKCSK LVG
121 ERRVIVPADP TLWTQEHVRQ WLEWAIKEYS LMEIDTSFFQ NMDGKELCKM NKEDFLRATT
   TQEHVRQ
181 LYNTEVLLSH LSYLRESSLL AYNTTSHTDQ SSRLSVKEDP SYDSVRGAW GNNMNSGLNK
   SSRLSVKE
241 SPPLGGAQTI SKNTEQRQPQ DPYQILGPTS SRLANPGSGQ IQLWQFLLEL LDSSANASCI
301 TWEQTNGEFK MTDPEVARR WGERKSKPNM NYDKLSRALR YYYDKNIMTK VHGKRYAYKF
361 DFHGIQAQLQ PHPTESSMYK YPSDISYMPH YHAHQQKVNF VPPHPSSMPV TSSSPFAASS
421 QYWTSPTGGI YPNPNVPRHP NTHVPSHLGS YY
   P NTHVPSHL
Figure 26

A

B

C

D

-1.75 0 1.75
fold change

- - + + + +
- + - + + +
- - + - - +

- - + + + +
- + - + + +
- - + - - +

- - + + + +
- + - + + +
- - + - - +

2.036 fold 791 fold

PSA/GAPDH

Target Gene/GAPDH

PSA ERG ETV1
Figure 29
SiRNA knockdown of ETV1 in LnCaP
Figure 30
siRNA knockdown of ERG in VCAP
Figure 32
Transgenic Mice

PolyA

3' primer

ERG1

5' primer

ARR2PB

PolyA

3' primer

ERG2

5' primer

ARR2PB

PolyA

3' primer

ETV1

5' primer

ARR2PB
Figure 33B

ERG/GAPDH

- Benign
- HGPIN
- PCa
- PCa (RP)
- Cell Line
**Figure 35**

<table>
<thead>
<tr>
<th></th>
<th>TMRSS2</th>
<th>ERG</th>
<th>ETV1</th>
<th>ETV4</th>
</tr>
</thead>
<tbody>
<tr>
<td># Evaluable cases (of 75):</td>
<td>57 (64.9%)</td>
<td>65</td>
<td>53</td>
<td>58 (1.7%)</td>
</tr>
<tr>
<td># Rearranged (%):</td>
<td>37 (64.9%)</td>
<td>36 (55.4%)</td>
<td>1 (1.9%)</td>
<td>1 (1.7%)</td>
</tr>
<tr>
<td># Deletion (%):</td>
<td>15 (40.5%)</td>
<td>14 (38.9%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td># All 4 probes evaluable (of 75):</td>
<td>38 (65.8%)</td>
<td>38</td>
<td>38</td>
<td>38</td>
</tr>
<tr>
<td># Rearranged (%):</td>
<td>25 (65.8%)</td>
<td>22 (57.9%)</td>
<td>1 (2.6%)</td>
<td>1 (2.6%)</td>
</tr>
<tr>
<td># Deletion (%):</td>
<td>10 (40.0%)</td>
<td>9 (40.9%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Diagram B:

- Black: Rearrangement Negative
- Black with white: Rearrangement Positive
- Black with white and grey: Rearrangement Positive (Deletion)
<table>
<thead>
<tr>
<th>Gene</th>
<th>Exon</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMRPSS2</td>
<td>1</td>
<td>cgcgcgctaaacgagaagcagccggagcgcggagcgcgcggagccgctgggccgcag</td>
</tr>
<tr>
<td>TMRPSS2</td>
<td>1,2</td>
<td>cgcgcgctaaacgagaagcagccggagcgcggagcgcgcggagccgctgggccgcag</td>
</tr>
<tr>
<td>TMRPSS2</td>
<td>1,2,3</td>
<td>cgcgcgctaaacgagaagcagccggagcgcggagcgcgcggagccgctgggccgcag</td>
</tr>
<tr>
<td>TMRPSS2</td>
<td>1,2,3,4</td>
<td>cgcgcgctaaacgagaagcagccggagcgcggagcgcgcggagccgctgggccgcag</td>
</tr>
<tr>
<td>TMRPSS2</td>
<td>1,2,3,4,5</td>
<td>cgcgcgctaaacgagaagcagccggagcgcggagcgcgcggagccgctgggccgcag</td>
</tr>
<tr>
<td>TMRPSS2</td>
<td>upstream long</td>
<td>CTTTGATAAAATAGTTTGAAGGAGGCCAGCTCGATTAAGACGCCTTTCGCCCTCCCTGCGAC</td>
</tr>
<tr>
<td>TMRPSS2</td>
<td>upstream short</td>
<td>ATCGTAAAGAGCCTTCTCCCGCTTCCCGAG</td>
</tr>
<tr>
<td>ERG</td>
<td>2</td>
<td>gttatttcaggcttctgagagccgagcagcgtgagcgcacaaagcacagacaaatgactca cagagaaaaagaggtgacgcagaaacaggggcaactcaag</td>
</tr>
<tr>
<td>ERG</td>
<td>3</td>
<td>ccctcagcaggtgcagactgcagagcgtggctacggagaccactcagctgccgaag</td>
</tr>
<tr>
<td>ERG</td>
<td>4</td>
<td>gactgcttaactgctgcagagcgtggctacggagaccactcagctgccgaag</td>
</tr>
<tr>
<td>ERG</td>
<td>5</td>
<td>gactgcttaactgctgcagagcgtggctacggagaccactcagctgccgaag</td>
</tr>
<tr>
<td>ERG</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>aacctactagtctagcaccatgctggcgaagaggggagaaaataagcc</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>cctctccattatgactagatgttatgataagctccctcagggaaagct</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>attaggagcccaccgagatccgcctgagcgggccagcggagccagc</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>taggcgtcaccgctcgcagggctacttcaccagtactactgag</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>gggagaagctccaggggctgctgcccaagtactactagtc</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>atgctccagagctccagagctgcagcctggcagcagctggcagc</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>atccctagtccagagctccagagctgcagcctggcagcagctggcagc</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>gggagaagctccaggggctgctgcccaagtactactagtc</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>ctgctccagagctccagagctgcagcctggcagcagctggcagc</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>atgctccagagctccagagctgcagcctggcagcagctggcagc</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>gggagaagctccaggggctgctgcccaagtactactagtc</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ETV1</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ttcctcctgtctagcaccatgagacgagtgtagcttctgccagcagctagc</td>
</tr>
<tr>
<td>2</td>
<td>ttgctccctgagccatagtcagtcagtcagtcagtcagtcagtcagtcagtc</td>
</tr>
<tr>
<td>3</td>
<td>ttgctccctgagccatagtcagtcagtcagtcagtcagtcagtcagtcagtc</td>
</tr>
<tr>
<td>4</td>
<td>ttgctccctgagccatagtcagtcagtcagtcagtcagtcagtcagtcagtc</td>
</tr>
<tr>
<td>5</td>
<td>ttgctccctgagccatagtcagtcagtcagtcagtcagtcagtcagtcagtc</td>
</tr>
<tr>
<td>6</td>
<td>ttgctccctgagccatagtcagtcagtcagtcagtcagtcagtcagtcagtc</td>
</tr>
<tr>
<td>7</td>
<td>ttgctccctgagccatagtcagtcagtcagtcagtcagtcagtcagtcagtc</td>
</tr>
<tr>
<td>8</td>
<td>ttgctccctgagccatagtcagtcagtcagtcagtcagtcagtcagtcagtc</td>
</tr>
<tr>
<td>9</td>
<td>ttgctccctgagccatagtcagtcagtcagtcagtcagtcagtcagtcagtc</td>
</tr>
<tr>
<td>10</td>
<td>ttgctccctgagccatagtcagtcagtcagtcagtcagtcagtcagtcagtc</td>
</tr>
<tr>
<td>11</td>
<td>ttgctccctgagccatagtcagtcagtcagtcagtcagtcagtcagtcagtc</td>
</tr>
<tr>
<td>Name</td>
<td>Reference</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------</td>
</tr>
<tr>
<td>TMPRSS2:ETV1a</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>TMPRSS2:ETV1b</td>
<td>1, 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>TMPRSS2 Exon</th>
<th>ERG Exon</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMPRSS2:ERG(1,2)</td>
<td>TMPRSS2:ERGb</td>
<td>1</td>
<td>2-11</td>
</tr>
<tr>
<td>TMPRSS2:ERG(1,3)</td>
<td>3</td>
<td>1</td>
<td>3-11</td>
</tr>
<tr>
<td>TMPRSS2:ERG(1,4)</td>
<td>TMPRSS2:ERGa</td>
<td>1</td>
<td>4-11</td>
</tr>
<tr>
<td>TMPRSS2:ERG(1,5)</td>
<td>2</td>
<td>1</td>
<td>5-11</td>
</tr>
<tr>
<td>TMPRSS2:ERG(2,2)</td>
<td>3</td>
<td>1, 2</td>
<td>2-11</td>
</tr>
<tr>
<td>TMPRSS2:ERG(2,3)</td>
<td>1, 2</td>
<td>1, 2</td>
<td>3-11</td>
</tr>
<tr>
<td>TMPRSS2:ERG(2,4)</td>
<td>1</td>
<td>1, 2</td>
<td>4-11</td>
</tr>
<tr>
<td>TMPRSS2:ERG(2,5)</td>
<td>2</td>
<td>1, 2</td>
<td>5-11</td>
</tr>
<tr>
<td>TMPRSS2:ERG(3,4)</td>
<td>3</td>
<td>1, 3</td>
<td>4-11</td>
</tr>
<tr>
<td>TMPRSS2:ERG(4,4)</td>
<td>1</td>
<td>1, 4</td>
<td>4-11</td>
</tr>
<tr>
<td>TMPRSS2:ERG(4,5)</td>
<td>1</td>
<td>1-4</td>
<td>5-11</td>
</tr>
<tr>
<td>TMPRSS2:ERG(5,4)</td>
<td>1</td>
<td>1-5</td>
<td>4-11</td>
</tr>
<tr>
<td>Gene</td>
<td>Accession number</td>
<td>TMPRSS2:ETV4a</td>
<td>TMPRSS2:ETV4b</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------</td>
<td>-----------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>TMPRSS2</td>
<td>NM_005656.2</td>
<td>upstream long</td>
<td>intron 2, exons 3-13</td>
</tr>
<tr>
<td>ERG</td>
<td>NM_004449.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETV1</td>
<td>NM_004956.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETV4</td>
<td>NM_001986.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMPRSS2:ERGa</td>
<td>DQ204772</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMPRSS2:ERGb</td>
<td>DQ204773.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMPRSS2:ETV1a</td>
<td>DQ204770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMPRSS2:ETV1b</td>
<td>DQ204771</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMPRSS2:ETV4a</td>
<td>DQ396625</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMPRSS2:ETV4b</td>
<td>DQ396626</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>References</th>
<th>PMID</th>
<th>Gene</th>
<th>Chromosomes Cancer 2006 Jul;45(7):717-9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16575875</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>16951141</td>
<td>Cancer Res. 2006 Sep 1,65(17):8347-51</td>
<td></td>
</tr>
</tbody>
</table>

Figure 36 (cont)
### Figure 37

<table>
<thead>
<tr>
<th>Primer</th>
<th>SEQ ID NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL11_exon1-f</td>
<td>GCTGCAAGACCTTGGCACAATGAC</td>
</tr>
<tr>
<td>FL11_exon2-r</td>
<td>TCACACAGGACAGCAGCTCCTTA</td>
</tr>
<tr>
<td>FL11_exon2/3-f</td>
<td>ACCACATGAAGGATCCAGGGAGTCT</td>
</tr>
<tr>
<td>FL11_exon3-r</td>
<td>ACCAGCTTGCTGCTATTGCTAAGC</td>
</tr>
<tr>
<td>FL11_exon4-f</td>
<td>CTAGCCGACCAACCCCTCTA</td>
</tr>
<tr>
<td>FL11_exon45-r</td>
<td>GTCACAGGAATGAAAATTGCTGAG</td>
</tr>
<tr>
<td>FL11_exon3a-f</td>
<td>CCTCTCAAGACATGACACACA</td>
</tr>
<tr>
<td>FL11_exon3a-r</td>
<td>GTCCTCGGGGAGCGATAGACTTC</td>
</tr>
<tr>
<td>FL11_exon5-f</td>
<td>GATGTAAGGCATGCTTCTGAGC</td>
</tr>
<tr>
<td>FL11_exon5-r</td>
<td>CTCACGTTCGAAAAGCAAGCTG</td>
</tr>
<tr>
<td>FL11_exon7-f</td>
<td>CTTGGAGGGGACCAAACGAT</td>
</tr>
<tr>
<td>FL11_exon8-r</td>
<td>GTCGCGGCCCAGATGCTGAC</td>
</tr>
<tr>
<td>FL11_exon9-f</td>
<td>GTCGACGGCGAGTGTACCA</td>
</tr>
<tr>
<td>FL11_exon9-r</td>
<td>AGCGGGCCGAGCCTAGC</td>
</tr>
<tr>
<td>FL11_exon10-f</td>
<td>ATGTITTATGACAAAGACGATTCTTGTGTC</td>
</tr>
<tr>
<td>FL11_exon10-r</td>
<td>ATGACGGGTAAGTCCCATGATTCTG</td>
</tr>
</tbody>
</table>

**ERG**
- forward- | CGAAGGCTGCTCAACCATCTC | 218 |
- reverse- | TAACTGAGGACGCTGCTTCA | 219 |
- FAM labeled MGB probe | CCACAGTGGCCAAAAA | 220 |

**FL11**
- forward- | CGGCAAAAGATATGCTTTACAAATT | 222 |
- reverse- | GACGACTCGGTGCTGATGT | 223 |
- FAM labeled MGB probe | CACGCGCATTGCCCA | 224 |

**TMPRSS2:ERG**
- specific
  - forward- | CGCGGCGAGGAAGCCTTA | 225 |
  - reverse- | TCCGTAGGACACTCAAATAAC | 226 |
  - FAM labeled MGB probe | CAGTGTGAGTGAGGACC | 227 |

**ETV4**
- forward- | GCTCGCTCCGATATATATTGAGAA | 228 |
- reverse- | CACACACAAACTTTGTAACGTAACG | 229 |
- FAM labeled MGB probe | ACCAGCCACTTCTGC | 230 |

**ETV1**
- forward- | GGTCGAGGATGGAATTTAATTGATTCA | 231 |
- reverse- | GCTGCGCTGTTTTCTGATG | 232 |
- FAM labeled MGB probe | TCGGCGACCCTTCTC | 233 |
Recurrent Gene Fusions in Prostate Cancer

Cross Reference to Related Applications


Statement Regarding Federally Sponsored Research or Development

This invention was made with government support under Grant Nos.: prostate SPORE P50CA69568, RO1 CA97063, U01 CA111275, CA046592 and AG021404 awarded by National Institutes of Health, and Grant No.: ARMY W81XWH-05-1-0173 awarded by the Department of Defense. The government has certain rights in the invention.

Field of the Invention

The present invention relates to compositions and methods for cancer diagnosis, research and therapy, including but not limited to, cancer markers. In particular, the present invention relates to recurrent gene fusions as diagnostic markers and clinical targets for prostate cancer.

Background of the Invention

A central aim in cancer research is to identify altered genes that are causally implicated in oncogenesis. Several types of somatic mutations have been identified including base substitutions, insertions, deletions, translocations, and chromosomal gains and losses, all of which result in altered activity of an oncogene or tumor suppressor gene. First hypothesized in the early 1990’s, there is now compelling evidence for a causal role for chromosomal rearrangements in cancer (Rowley, Nat Rev Cancer 1: 245 (2001)). Recurrent chromosomal aberrations were thought to be primarily characteristic of leukemias, lymphomas, and sarcomas. Epithelial tumors (carcinomas), which are much more common and contribute to a relatively large fraction of the morbidity and mortality associated with human cancer, comprise less than 1% of the known, disease-specific chromosomal rearrangements (Miletman, Mutat Res 462: 247 (2000)). While hematological malignancies are often characterized by balanced, disease-specific chromosomal rearrangements, most solid tumors have a plethora of non-specific chromosomal aberrations. It is thought that the karyotypic complexity of solid tumors is due to secondary alterations acquired through cancer evolution or progression.

Two primary mechanisms of chromosomal rearrangements have been described. In one mechanism, promoter/enhancer elements of one gene are rearranged adjacent to a proto-oncogene, thus causing altered expression of an oncogenic protein. This type of translocation is exemplified by the apposition of immunoglobulin (Ig) and T-cell receptor (TCR) genes to MYC leading to activation of this oncogene in B- and T-cell malignancies, respectively (Rabbits, Nature 372: 143 (1994)). In the second mechanism, rearrangement results in the fusion of two genes, which produces a fusion protein that may have a new function or altered activity. The prototypic example of this translocation is the BCR-ABL gene fusion in chronic myelogenous leukemia (CML) (Rowley, Nature 243: 290 (1973); de Klein et al., Nature 300: 765 (1982)). Importantly, this finding led to the rational development of imatinib mesylate (Gleevec), which successfully targets the BCR-ABL kinase (Deininger et al., Blood 105: 2640 (2005)). Thus, identifying recurrent gene rearrangements in common epithelial tumors may have profound implications for cancer drug discovery efforts as well as patient treatment.

Summary of the Invention

The present invention provides, but is not limited to, methods for diagnosing prostate cancer in a patient comprising: providing a sample from the patient; and, detecting the presence or absence in the sample of a gene fusion having a 5’ portion from a transcriptional regulatory region of an androgen regulated gene (ARG) and a 3’ portion from an ETS family member gene, wherein the presence in the sample of the gene fusion is indicative of prostate cancer in the patient. The ARG may be TMPRSS2 or PSA. The ETS family member gene may be ARG, ET1V1 (ER81), EL1, ETS1, ETS2, ELK1, ETV6 (TEL1), ETV7 (TEL2), GABPB1, EL1F, ETV4 (E1A1F, PEA3), ETV5 (ER3), ERF, PEA3/E1A1F, PU1, ESE1/EFS, SAP1 (ELK4), ETV3 (MET), EWS/FL11, ESE1, ESE2 (ELF5), ESE3, PDE, NET (ELK3; SAP2), NERF (ELF2), or FEV. The transcriptional regulatory region of the ARG may comprise a promoter region of the ARG. The promoter region of the ARG may further comprise an androgen response element (ARE) of the ARG.

Detecting the presence or absence in the sample of a gene fusion may comprise detecting chromosomal rearrangements of genomic DNA having a 5’ portion from a transcriptional regulatory region of an ARG and a 3’ portion from an ETS family member gene. A variety of techniques may be used for detecting the chromosomal rearrangements of genomic DNA, including nucleic acid sequencing, nucleic acid hybridization, and, nucleic acid amplification. Nucleic acid hybridization techniques include in situ hybridization (ISH), microarray, and Southern blot. Nucleic acid amplification techniques include polymerase chain reaction (PCR), reverse transcription polymerase chain reaction (RT-PCR), transcription-mediated amplification (TMA), ligase chain reaction (LCR), strand displacement amplification (SDA), and nucleic acid sequence based amplification (NASBA).

Detecting the presence or absence in the sample of a gene fusion may alternatively comprise detecting chimeric mRNA transcripts having a 5’ portion from a transcriptional regulatory region of an ARG and a 3’ portion from an ETS family member gene. A variety of techniques may be used for detecting the chimeric mRNA, including nucleic acid sequencing, nucleic acid hybridization, and, nucleic acid amplification. Nucleic acid hybridization techniques include in situ hybridization (ISH) (e.g., Fluorescence in situ hybridization (FISH)), microarray, and Northern blot. Nucleic acid amplification techniques include polymerase chain reaction (PCR), reverse transcription polymerase chain reaction (RT-PCR), transcription-mediated amplification (TMA), ligase chain reaction (LCR), strand displacement amplification (SDA), and nucleic acid sequence based amplification (NASBA).

Detecting the presence or absence in the sample of a gene fusion may also alternatively comprise detecting an amino-terminally truncated ETS family member protein resulting from a fusion of a transcriptional regulatory region of an ARG to an ETS family member gene, or detecting a chimeric protein having an amino-terminal portion from a transcriptional
regulatory region of an ARG and a carboxy-terminal portion from an ETS family member gene. A variety of techniques may be used for detecting the truncated ETS family member protein or chimeric protein: protein sequencing; and, immunoassay. Immunoassay techniques include immunoprecipitation, Western blot, ELISA, immunohistochemistry, immunocytochemistry, flow cytometry, and immuno-PCR.

The present invention further provides, but is not limited to, compositions and kits for diagnosing prostate cancer in a patient. The compositions and kits may comprise: a single labeled probe comprising a sequence that hybridizes to the junction at which a 5' portion from a transcriptional regulatory region of an ARG fuses to a 3' portion from an ETS family member gene; a pair of labeled probes wherein the first labeled probe comprises a sequence that hybridizes to a transcriptional regulatory region of an ARG and the second labeled probe comprises a sequence that hybridizes to an ETS family member gene; a pair of amplification oligonucleotides wherein the first amplification oligonucleotide comprises a sequence that hybridizes to a transcriptional regulatory region of an ARG and the second amplification oligonucleotide comprises a sequence that hybridizes to an ETS family member gene; an antibody to an amino-terminally truncated ETS family member protein resulting from a fusion of a transcriptional regulatory region of an ARG to an ETS family member gene; or, an antibody to a chimeric protein having an amino-terminal portion from a transcriptional regulatory region of an ARG and a carboxy-terminal portion from an ETS family member gene.

The present invention also provides, but is not limited to, methods for treating prostate cancer in a patient comprising: administering to the patient an agent that inhibits at least one biological activity of a gene fusion having a 5' portion from a transcriptional regulatory region of an androgen regulated gene (ARG) and a 3' portion from an ETS family member gene. The ARG may be TMPRSS2 or PSA. The ETS family member gene may be ERG, ET1V (ER81), FLI1, ETN1, ET52, ELK1, ET6V (TEL1), ET7V (TEL2), GABPa, ELF1, ET4V (ELAF; PEAS), ET3V (ERMS), ERF, PEAS1/ELAF, PU1, ESE1/ESX, SAP1 (ELK4), ET3V (METS), EWS/FLI1, ESE1, ESE2 (ELF5), ESE3, PDEF, NET (ELK3; SAP2), NERF (ELF2), and FEV. The transcriptional regulatory region of the ARG may comprise a promoter region of the ARG. The promoter region of the ARG may further comprise an androgen response element (ARE) of the ARG. The agent may be a small molecule, an siRNA, an antisense nucleic acid, or an antibody.

Additional embodiments of the present invention are provided in the description and examples below.

DESCRIPTION OF THE FIGURES

FIG. 1 shows the Cancer Outlier Profile Analysis (COPA) of microarray data. (A) ET1V (left panels) and ERG (middle panels) expression (normalized expression units) are shown from all profiles samples in two large scale gene expression studies. (B) As in (A), except data from laser capture micro-dissected samples were used. (C) As in (A), except oncogenes (FGFR3 and CCND1) with known translocations to the immunoglobulin heavy chain promoter (IGH) in multiple myeloma were examined.

Fig. 2 shows the identification and characterization of TMPRSS2:ET1V and TMPRSS2:ERG gene fusions in prostate cancer (PCA). (A) Prostate cancer cell lines (DuCaP, LnCaP and VCaP) and hormone refractory metastatic (MET) prostate cancer tissues were analyzed for ERG (I) and ET1V (C) mRNA expression by quantitative PCR (QPCR). (B) Loss of over-expression of ET1V exons 2 and 3 in MET26 compared to LnCaP cells. (C) Schematic of 5' RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE) results for ET1V in MET26-LN and ERG in MET28-LN revealing gene fusions with TMPRSS2. (D) Validation of TMPRSS2:ET1V expression using translocation-specific QPCR in MET26-LN and MET26-RP. (E) Validation of TMPRSS2:ERG expression using translocation-specific QPCR in cell lines and PCA specimens.

FIG. 3 shows interphase fluorescence in situ hybridization (FISH) on formalin-fixed paraffin embedded tissue sections that confirms TMPRSS2:ET1V gene fusion and ERG gene rearrangement. (A and B) Show two-color, fusion-signal approach to detect the fusion of TMPRSS2 (green signal) and ET1V (red signal). (C and D) Detection of ERG gene rearrangements using a two-color split-signal approach with two probes spanning the 5' (green signal) and 3' (red signal) regions of ERG. (E) Matrix representation of FISH results using the same probes as (A-D) on an independent tissue microarray containing cores from 13 cases of clinically localized prostate cancer (PCA) and 16 cases of metastatic prostate cancer (MET).

FIG. 4 shows androgen regulation of ERG in prostate cancer cells carrying the TMPRSS2:ERG translocation.

FIG. 5 shows Cancer Outlier Profile Analysis (COPA). FIG. 5A shows a schematic of COPA analysis. FIG. 5B shows that RUNX1T1 (ETO) had the highest scoring outlier profile at the 90th percentile in the Valk et al. acute myeloid leukemia dataset (n=293).

FIG. 6 shows a schematic of RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE) results for ET1V in MET26-LN and ERG in PCA4 revealing gene fusions with TMPRSS2 (TMPRSS2:ERG fusion).

FIG. 7 shows over-expression of ETS family members in prostate cancer. Expression of all monitored ETS family members in profiled benign prostate, prostatic intraepithelial neoplasia (PIN), clinically localized prostate cancer and metastatic prostate cancer from grossly dissected tissue (A) or tissue isolated by laser capture microdissection (B) was visualized using Oncomine.

FIG. 8 shows over expression of TMPRSS2 and ET4V loci in a prostate cancer case that over-expresses ET4V. A. Expression of the indicated exons or region of ET4V in pooled benign prostate tissue (CPP), prostate cancers that did not over-express ET4V and were either TMPRSS2:ERG positive (PCA1-2) or negative (PCA5-4), and the prostate cancer case from our LCM cohort with ET4V over-expression (PCA5). B. RLM-RACE reveals fusion of sequences upstream of TMPRSS2 with ET4V in PCA5. C. Expression of TMPRSS2:ET4Va and TMPRSS2:ET4Vb in PCA5 by QPCR. D. Interphase fluorescence in situ hybridization on formalin-fixed paraffin-embedded tissue confirms fusion of TMPRSS2 and ET4V loci in PCA5.

FIG. 9 (SEQ ID NOS: 32-54) shows mRNA sequences of exemplary ETS family genes.

FIG. 10 (SEQ ID NO.: 307) shows the mRNA sequence of TMPRSS2.

FIG. 11 shows TMPRSS2:ERG gene fusion analysis by Fisher. Panel A: Ideogram, depicting a break apart assay for the indirect detection of TMPRSS2:ERG fusion. Panel B: Interphase nuclei of a stromal cell (left) and a prostate cancer gland (right). Panel C: Interphase nuclei of prostate cancer glands showing break apart and simultaneous deletion as indicated by loss of the telomeric probe (100x oil immersion objective magnification). Panel D: Magnified view of boxed
area in C demonstrating two nuclei with break apart and loss of the telomeric probe. (60x oil immersion objective magnification).

FIG. 12 shows Genomic deletions on chromosome 21 between ERG and TMPRSS2. Panel A: Samples, including 6 cell lines, 13 xenografts and 11 metastatic PCA samples, were characterized for TMPRSS2:ERG and TMPRSS2:ETV1 status (gray bars for negative and blue bar for positive status), by qPCR and/or by FISH. Panel B: Magnification of the green framed box in A. Panel C: Magnification of the black framed box in A.

FIG. 13 shows TMPRSS2:ERG rearrangement in clinically localized prostate cancer and association with pathological parameters. Panel A. The TMPRSS2:ERG rearrangement was identified in 49.2% of the primary PCA samples and 41.2% in the hormone naïve metastatic LN samples. Panel B. TMPRSS2:ERG rearranged tumors with deletions tended to be observed in a higher percentage of PCA cases with advanced tumor stage (p=0.03).

FIG. 14 shows known genes located 21q22-23 between ERG (centromeric) and TMPRSS2 (telomeric). Genes above the black line are oriented 5'-centromeric to 3'-telomeric and genes below the black line are oriented 5'-telomeric to 3'-centromeric. In the lower half of the image, a magnification of the ERG locus is depicted with FISH probes.

FIG. 15 shows 'heterogenous' prostate cancer case predominantly showing TMPRSS2:ERG rearrangement with the deletion (nucleus on the right) and only small areas showing the TMPRSS2:ERG rearrangement without the deletion (nucleus on the left).

FIG. 16 shows meta-analysis of genes located between TMPRSS2 and ERG across 8 published expression array datasets.

FIG. 17 shows that the FISH assay detects the characteristic deletion associated with TMPRSS2:ERG gene fusion, which is associated with disease progression. Panels A and B: For analyzing the ERG rearrangement on chromosome 21q22.2, a break apart probe system was applied, consisting of the Biotin-14-dUTP labeled BAC clone RP11-24A11 (eventually conjugated to produce a red signal) and the Digoxigenin-dUTP labeled BAC clone RP11-137J13 (eventually conjugated to produce a green signal), spanning the neighboring centromeric and telomeric region of the ERG locus, respectively. Using this break apart probe system, a nucleus without ERG rearrangement exhibits two pairs of juxtaposed red and green signals. Juxtaposed red-green signals form a yellow fusion signal (Panel B, arrow). Panel C: In a cumulative incidence regression model, TMPRSS2:ERG was evaluated as a determinant for the cumulative incidence of metastases or prostate cancer-specific death.

FIG. 18 shows FLI1 overexpression without fusion transcript.

FIG. 19 shows induction of ERG protein expression by androgen in TMPRSS2:ERG+ cells.

FIG. 20 (SEQ ID NOS: 236-239) shows a schematic of the endogenous and fusion ERG polypeptides.

FIG. 21 shows Nuclear interactors for ERG2.

FIG. 22 (SEQ ID NOS: 240-241) shows sequences for peptide antibody and aqua probe generation against ERG1.

FIG. 23 (SEQ ID NOS: 242-245) shows sequences for peptide antibody and aqua probe generation against ETV1.

FIG. 24 (SEQ ID NOS: 246-252) shows sequences for peptide antibody and aqua probe generation against FLI1.

FIG. 25 (SEQ ID NOS: 253-268) shows sequences for peptide antibody and aqua probe generation against ETV4.

FIG. 26 shows the over-expression and androgen regulation of ETV1 in the LNCaP prostate cancer cell line. FIG. 26A shows expression signature of androgen-regulated genes in VCaP and LNCaP prostate cancer cell lines. FIG. 26B shows confirmation of PSA induction by androgen in both VCaP and LNCaP cells by quantitative PCR (qPCR).

FIG. 26C shows ETV1 induction by androgen in LNCaP cells. FIG. 26D shows that ETV1 is markedly over-expressed in LNCaP cells.

FIG. 27 shows rearrangement of ETV1 in LNCaP cells. FIG. 27A shows a schematic of BACs used as probes for fluorescence in situ hybridization (FISH). FIG. 27B shows that RP11-124L22 and RP11-114G13 co-localize to chromosome 7 in normal peripheral lymphocytes (NPLs). FIG. 27C shows localization of BAC #1 and BAC #4 on metaphase spreads (top panel) and interphase cells (bottom panel) was determined in the near tetraploid LNCaP cell line. FIG. 27D shows signal from RP11-124L22 localizes to chromosome 14 in LNCaP cells.

FIG. 28 shows that the entire ETV1 locus is inserted into chromosome 14 in LNCaP cells. FIG. 28A shows a schematic of BACs used in this experiment. FIG. 28B shows localization of RP11-124L22 (BAC #1) and RP11-313C20 (BAC #2) on metaphase spreads (top panel) and interphase cells (bottom panel) was determined by FISH in LNCaP cells.

FIG. 29 shows siRNA knockdown of ETV1 in LNCaP.

FIG. 30 shows siRNA knockdown of ETV1 in VCaP.

FIG. 31 shows viral overexpression systems.

FIG. 32 shows a schematic of transgenic mice.

FIG. 33 shows detection of ERG and ETV1 transcripts in urine. FIG. 33A shows detection of ERG and ETV1 in LNCaP (high ETV1 expression) or VCaP (high ERG and TMPRSS2:ERG expression) prostate cancer cells. FIG. 33B shows detection of ERG and ETV1 in urine of patients suspected of having prostate cancer.

FIG. 34 shows assays used to detect TMPRSS2:ETS gene fusions in prostate cancer. FIG. 34A shows break apart assays for TMPRSS2 and ERG. An ERG rearrangement positive case (without deletion), as indicated by one pair of split 5' and 3' signals, is shown in the left panel. A TMPRSS2 rearrangement positive case (with deletion), as indicated by a loss of one 3' signal, is shown in the right panel. FIG. 34B shows a fusion assay for TMPRSS2:ETV1 gene fusions. FIG. 34C shows a break apart assay for ETV4.

FIG. 35 shows TMPRSS2, ERG, ETV1 and ETV4 rearrangements as detected by FISH. FIG. 35A shows a Table of results for rearrangements in TMPRSS2, ERG, ETV1 and ETV4 as detected by the assays shown in FIG. 34. FIG. 35B shows a heat map representation of the TMPRSS2, ERG, ETV1 and ETV4 status from the 38 cases where all four assays were evaluable as described in A. FIG. 35C shows a heat map representation of cases with discordant TMPRSS2 and ETS rearrangement status.

FIG. 36 (SEQ ID NOS: 269-306) shows the sequences of gene fusions of the present invention.

FIG. 37 shows primers and probes for FLI1 expression analysis.

DEFINITIONS

To facilitate an understanding of the present invention, a number of terms and phrases are defined below:

As used herein, the term “gene fusion” refers to a chimeric genomic DNA, a chimeric messenger RNA, a truncated protein or a chimeric protein resulting from the fusion of at least a portion of a first gene to at least a portion of a second gene. The gene fusion need not include entire genes or exons of genes.
As used herein, the term “transcriptional regulatory region” refers to the non-coding upstream regulatory sequence of a gene, also called the 5’ untranslated region (5’ UTR).

As used herein, the term “androgen regulated gene” refers to a gene or portion of a gene whose expression is initiated or enhanced by an androgen (e.g., testosterone). The promoter region of an androgen regulated gene may contain an “androgen response element” that interacts with androgens or androgen signaling molecules (e.g., downstream signaling molecules).

As used herein, the terms “detect”, “detecting”, or “detection” may describe either the general act of discovering or discerning or the specific observation of a detectably labeled composition.

As used herein, the term “inhibits at least one biological activity of a gene fusion” refers to any agent that decreases any activity of a gene fusion of the present invention (e.g., including, but not limited to, the activities described herein), via directly contacting gene fusion protein, contacting gene fusion mRNA or genomic DNA, causing conformational changes of gene fusion polypeptides, decreasing gene fusion protein levels, or interfering with gene fusion interactions with signaling partners, and affecting the expression of gene fusion target genes. Inhibitors also include molecules that indirectly regulate gene fusion biological activity by intersecting upstream signaling molecules.

As used herein, the term “siRNAs” refers to small interfering RNAs. In some embodiments, siRNAs comprise a duplex, or double-stranded region, of about 18-25 nucleotides long; often siRNAs contain from about two to four unpaired nucleotides at the 3’ end of each strand. At least one strand of the duplex or double-stranded region of a siRNA is substantially homologous to, or substantially complementary to, a target RNA molecule. The strand complementary to a target RNA molecule is the “sense strand,” and is also complementary to the siRNA antisense strand. siRNAs may also contain additional sequences; non-limiting examples of such sequences include linking sequences, or loops, as well as stem and other folded structures. siRNAs appear to function as key intermediaries in triggering RNA interference in invertebrates and in vertebrates, and in triggering sequence-specific RNA degradation during posttranscriptional gene silencing in plants.

The term “RNA interference” or “RNAi” refers to the silencing or decreasing of gene expression by siRNAs. It is the process of sequence-specific, post-transcriptional gene silencing in animals and plants, initiated by siRNA that is homologous in its duplex region to the sequence of the silenced gene. The gene may be endogenous or exogenous to the organism, present integrated into a chromosome or present in a transfection vector that is not integrated into the genome. The expression of the gene is either completely or partially inhibited. RNAi may also be considered to inhibit the function of a target RNA; the function of the target RNA may be complete or partial.

As used herein, the term “stage of cancer” refers to a qualitative or quantitative assessment of the level of advancement of a cancer. Criteria used to determine the stage of a cancer include, but are not limited to, the size of the tumor and the extent of metastases (e.g., localized or distant).

As used herein, the term “gene transfer system” refers to any means of delivering a composition comprising a nucleic acid sequence to a cell or tissue. For example, gene transfer systems include, but are not limited to, vectors (e.g., retroviral, adeno viral, adeno-associated viral, and other nucleic acid-based delivery systems), microinjection of naked nucleic acid, polymer-based delivery systems (e.g., liposome-based and metallic particle-based systems), biolistic injection, and the like. As used herein, the term “viral gene transfer system” refers to gene transfer systems comprising viral elements (e.g., intact viruses, modified viruses and viral components such as nucleic acids or proteins) to facilitate delivery of the sample to a desired cell or tissue. As used herein, the term “adenovirus gene transfer system” refers to gene transfer systems comprising intact or altered viruses belonging to the family Adenoviridae.

As used herein, the term “site-specific recombination target sequences” refers to nucleic acid sequences that provide recognition sequences for recombination factors and the location where recombination takes place.

As used herein, the term “nucleic acid molecule” refers to any nucleic acid containing molecule, including but not limited to, DNA or RNA. The term encompasses sequences that include any of the known base analogs of DNA and RNA including, but not limited to, 4-acetylcytosine, 5-hydroxy-N-6-methyladenosine, azacytidine, pseudouridine, 5-(carboxyhydroxymethyl) uracil, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxyaminoethylaminomethyluracil, dicytosine, inosine, N6-isopentenyladenine, 1-methyladenine, 1-methyl pseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethyl guanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-methyladenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosyloligoceosine, 5′-methoxy carbonylmethyluracil, 5-methoxyuracil, 2-methylthio-N-6-isopentenyladenine, uracil-5-oxyacetic acid methyl ester, uracil-5-oxyacetic acid, oxybutoxosine, pseudouracil, queosine, 2-thiouracil, 5-methyl-2-thiouracil, 4-thiouracil, 5-methyluracil, N-uracil-5-oxyacetic acid methyl ester, uracil-5-oxyacetic acid, pseudouracil, queosine, 2-thiouracil, and 2,6-diaminopurine.

The term “gene” refers to a nucleic acid (e.g., DNA) sequence that comprises coding sequences necessary for the production of a polypeptide, precursor, or RNA (e.g., rRNA, tRNA). The polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired activity or functional properties (e.g., enzymatic activity, ligand binding, signal transduction, immunogenicity, etc.) of the full-length or fragment are retained. The term also encompasses the coding region of a structural gene and the sequences located adjacent to the coding region on both the 5′ and 3′ ends for a distance of about 1 kb or more on either end such that the gene corresponds to the length of the full-length mRNA. Sequences located 5′ of the coding region and present on the mRNA are referred to as 5′ non-translated sequences. Sequences located 3′ or downstream of the coding region and present on the mRNA are referred to as 3′ non-translated sequences. The term “gene” encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region uninterrupted with non-coding sequences termed “introns” or “intervening regions” or “intervening sequences.” Introns are segments of a gene that are transcribed into nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or “spliced out” from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide.

As used herein, the term “heterologous gene” refers to a gene that is not in its natural environment. For example,
heterologous gene includes a gene from one species introduced into another species. A heterologous gene also includes a gene native to an organism that has been altered in some way (e.g., mutated, added in multiple copies, linked to non-native regulatory sequences, etc). Heterologous genes are distinguished from endogenous genes in that the heterologous gene sequences are typically joined to DNA sequences that are not found naturally associated with the gene sequences in the chromosome or are associated with portions of the chromosome not found in nature (e.g., genes expressed in loci where the gene is not normally expressed).

As used herein, the term "oligonucleotide," refers to a short length of single-stranded polynucleotide chain. Oligonucleotides are typically less than 200 residues long (e.g., between 15 and 100), however, as used herein, the term is also intended to encompass longer polynucleotide chains. Oligonucleotides are often referred to by their length. For example a 24 residue oligonucleotide is referred to as a "24-mer". Oligonucleotides can form secondary and tertiary structures by self-hybridizing or by hybridizing to other polynucleotides. Such structures can include, but are not limited to, duplexes, hairpins, cruciforms, bends, and triplexes.

As used herein, the terms "complementary" or "complementarity" are used in reference to polynucleotides (i.e., a sequence of nucleotides) related by the base-pairing rules. For example, the sequence "5'-A-G-T-3'," is complementary to the sequence "3'-T-C-A-5'." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods that depend upon binding between nucleic acids.

The term "homology" refers to a degree of complementarity. There may be partial homology or complete homology (i.e., identity). A partially complementary sequence is a nucleic acid molecule that at least partially inhibits a completely complementary nucleic acid molecule from hybridizing to a target nucleic acid is "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous nucleic acid molecule to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target that is substantially non-complementary (e.g., less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe that can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described above.

A gene may produce multiple RNA species that are generated by differential splicing of the primary RNA transcript. cdNAS that are splice variants of the same gene will contain regions of sequence identity or complete homology (representing the presence of the same exon or portion of the same exon on both cdNAS) and regions of complete non-identity (for example, representing the presence of exon "A" on cdNA 1 wherein cdNA 2 contains exon "B" instead). Because the two cdNAS contain regions of sequence identity they will both hybridize to a probe derived from the entire gene or portions of the gene containing sequences found on both cdNAS; the two splice variants are therefore substantially homologous to such a probe and to each other.

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe that can hybridize (i.e., it is the complement of) the single-stranded nucleic acid sequence under conditions of low stringency as described above.

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the Tm of the formed hybrid, and the GC ratio within the nucleic acids. A single molecule that contains pairing of complementary nucleic acids within its structure is said to be "self-hybridized."

As used herein the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. Under "low stringency conditions" a nucleic acid sequence of interest will hybridize to its exact complement, sequences with single base mismatches, closely related sequences (e.g., sequences with 90% or greater homology), and sequences having only partial homology (e.g., sequences with 50-90% homology). Under "medium stringency conditions," a nucleic acid sequence of interest will hybridize only to its exact complement, sequences with single base mismatches, and closely related sequences (e.g., 90% or greater homology). Under "high stringency conditions," a nucleic acid sequence of interest will hybridize only to its exact complement, and (depending on conditions such a temperature) sequences with single base mismatches. In other words, under conditions of high stringency the temperature can be raised so as to exclude hybridization to sequences with single base mismatches.

"High stringency conditions" when used in reference to nucleic acid hybridization compromise conditions equivalent to binding or hybridization at 42° C in a solution consisting of 5×SSPE (43.8 g/l NaCl, 6.9 g/l NaH2PO4·H2O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5×Denhardt’s reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1×SSPE, 1.0% SDS at 42° C when a probe of about 500 nucleotides in length is employed.

"Medium stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42° C in a solution consisting of 5×SSPE (43.8 g/l NaCl, 6.9 g/l NaH2PO4·H2O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.2% SDS, 5×Denhardt’s reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1×SSPE, 1.0% SDS at 42° C when a probe of about 500 nucleotides in length is employed.

"Low stringency conditions" comprise conditions equivalent to binding or hybridization at 42° C in a solution consisting of 5×SSPE (43.8 g/l NaCl, 6.9 g/l NaH2PO4·H2O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5×Denhardt’s reagent [50×Denhardt’s contains per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V, Sigma)
and 100 μg/ml denatured salmon sperm DNA followed by washing in a solution comprising 5xSSPE, 0.1% SDS at 42°C. When a probe of about 500 nucleotides in length is employed.

The art knows well that numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, the art knows conditions that promote hybridization under conditions of high stringency (e.g., increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.) (see definition above for “stringency”).

As used herein, the term “amplification oligonucleotide” refers to an oligonucleotide that hybridizes to a target nucleic acid, or its complement, and participates in a nucleic acid amplification reaction. An example of an amplification oligonucleotide is a “primer” that hybridizes to a template nucleic acid and contains a 3' OH end that is extended by a polymerase in an amplification process. Another example of an amplification oligonucleotide is an oligonucleotide that is not extended by a polymerase (e.g., because it has a 3' blocked end) but participates in or facilitates amplification. Amplification oligonucleotides may optionally include modified nucleotides or analogs, or additional nucleotides that participate in an amplification reaction but are not complementary to or contained in the target nucleic acid. Amplification oligonucleotides may contain a sequence that is not complementary to the target or template sequence. For example, the 5' region of a primer may include a promoter sequence that is non-complementary to the target nucleic acid (referred to as a “promoter-primer”). Those skilled in the art will understand that an amplification oligonucleotide that functions as a primer may be modified to include a 5' promoter sequence, and thus function as a promoter-primer. Similarly, a promoter-primer may be modified by removal of, or synthesis without, a promoter sequence and still function as a primer. A 3' blocked amplification oligonucleotide may provide a promoter sequence and serve as a template for polymerization (referred to as a “promoter-provider”).

As used herein, the term “primer” refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, that is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product that is complementary to a nucleic acid strand is induced, (i.e., in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

As used herein, the term “probe” refers to an oligonucleotide (i.e., a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, that is capable of hybridizing to at least a portion of another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular gene sequences. It is contemplated that any probe used in the present invention will be labeled with any “reporter molecule,” so that is detectable in any detection system, including, but not limited to enzyme (e.g., ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

The term “isolated” when used in relation to a nucleic acid, as in “an isolated oligonucleotide” or “isolated polynucleotide” refers to a nucleic acid sequence that is identified and separated from at least one component or contaminant with which it is ordinarily associated in its natural source. Isolated nucleic acid is such present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids as nucleic acids such as DNA and RNA found in the state they exist in nature. For example, a given DNA sequence (e.g., a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs that encode a multitude of proteins. However, isolated nucleic acid encoding a given protein includes, by way of example, such nucleic acid in cells ordinarily expressing the given protein where the nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid, oligonucleotide, or polynucleotide may be present in single-stranded or double-stranded form. When an isolated nucleic acid, oligonucleotide or polynucleotide is to be utilized to express a protein, the oligonucleotide or polynucleotide will contain at a minimum the sense or coding strand (i.e., the oligonucleotide or polynucleotide may be single-stranded), but may contain both the sense and anti-sense strands (i.e., the oligonucleotide or polynucleotide may be double-stranded).

As used herein, the term “purified” or “to purify” refers to the removal of components (e.g., contaminants) from a sample. For example, antibodies are purified by removal of contaminating non-immunoglobulin proteins; they are also purified by the removal of immunoglobulin that does not bind to the target molecule. The removal of non-immunoglobulin proteins and/or the removal of immunoglobulins that do not bind to the target molecule results in an increase in the percent of target-reactive immunoglobulins in the sample. In another example, recombinant polypeptides are expressed in bacterial host cells and the polypeptides are purified by the removal of host cell proteins; the percent of recombinant polypeptides is thereby increased in the sample.

**DETAILED DESCRIPTION OF THE INVENTION**

The present invention is based on the discovery of recurrent gene fusions in prostate cancer. The present invention provides diagnostic, research, and therapeutic methods that either directly or indirectly detect or target the gene fusions. The present invention also provides compositions for diagnostic, research, and therapeutic purposes.

I. Gene Fusions

The present invention identifies recurrent gene fusions indicative of prostate cancer. The gene fusions are the result of a chromosomal rearrangement of an androgen regulated gene (ARG) and an ETS family member gene. Despite their recurrence, the junction where the ARG fuses to the ETS
family member gene varies. The gene fusions typically comprise a 5’ portion from a transcriptional regulatory region of an ARG and a 3’ portion from an ETS family member gene. The recurrent gene fusions have use as diagnostic markers and clinical targets for prostate cancer.

A. Androgen Regulated Genes

Genes regulated by androgenic hormones are of critical importance for the normal physiological function of the human prostate gland. They also contribute to the development and progression of prostate carcinoma. Recognized ARGs include, but are not limited to: TMPRSS2; PSA; PSMA; KLK2; SNRK; Seladin-1; and, FKBP51 (Paoloni-Giacobino et al., Genomics 44: 309 (1997); Velasco et al., Endocrinology 145(8): 3913 (2004)). TMPRSS2 (NM_0005656), in particular, has been demonstrated to be highly expressed in prostate epithelium relative to other normal human tissues (Lin et al., Cancer Research 59: 4180 (1999)). The TMPRSS2 gene is located on chromosome 21. This gene is located at 41,750,397-41,801,948 bp from the pter (51,151 total bp; minus strand orientation). The human TMPRSS2 protein sequence may be found at GenBank accession no. AAC51784 (Swiss Protein accession no. O15393) and the corresponding cDNA at GenBank accession no. U75329 (see also, Paoloni-Giacobino et al., Genomics 44: 309 (1997)).

The transcriptional regulatory region of an ARG may contain coding or non-coding regions of the ARG, including the promoter region. The promoter region of the ARG may further contain an androgen response element (ARE) of the ARG. The promoter region for TMPRSS2, in particular, is provided by GenBank accession number AJ276404.

B. ETS Family Member Genes

The ETS family of transcription factors regulate the intracellular signaling pathways controlling gene expression. As downstream effectors, they activate or repress specific target genes. As upstream effectors, they are responsible for the spatial and temporal expression of numerous growth factor receptors. Almost 30 members of this family have been identified and implicated in a wide range of physiological and pathological processes. These include, but are not limited to: ERG; ETV1 (ER81); FLI1; ETS1; ETS2; ELK1; ETV6 (TEL1); ETV7 (TEL2); GABPα; ELF1; ETV4 (EIAF; PEA3); ETV5 (ERM); ERF; PEA3/EIAF; PU.1; ESE1/ESX; SAP1 (ELK4); ETV3 (METS); EWS/FLI1; ESE1; ESE2 (ELFS); ESE3; PDEF; NET (ELK3; SAP2); NERF (ELF2); and FEV. Exemplary ETS family member gene sequences are given in FIG. 9.

ERG (NM_0004449), in particular, has been demonstrated to be highly expressed in prostate epithelium relative to other normal human tissues. The ERG gene is located on chromosome 21. The gene is located at 38,675,671-38,955,488 base pairs from the pter. The ERG gene is 279,817 total bp; minus strand orientation. The corresponding ERG cDNA and protein sequences are given at GenBank accession no. NM_001986 and GenBank accession no. NP_019777 (Swiss protein acc. no. P43268), respectively.

C. ARG/ETS Gene Fusions

As described above, the present invention provides fusions of an ARG to an ETS family member gene. Exemplary gene fusion sequences are given in FIG. 36. For all involved genes (TMPRSS2, ERG, ETV1 and ETV4), the GenBank reference sequence ID’s are provided and the exons are aligned using the May 2004 assembly of the UCSC Human Genome. For all identified fusions, FIG. 36 provides a complete sequence from the beginning of the TMPRSS2 gene through the fusion and the stop codon of the ETS family member gene. The deposited GenBank sequence for each of the published variants is also provided. Some TMPRSS2:ERG and TMPRSS2:ETV1 fusions are described by the breakpoint exons of TMPRSS2 and the ETS family member gene. For example, TMPRSS2:ERG, which fuses exon 1 of TMPRSS2 to exons 4 through 11 of ERG, is identified as TMPRSS2:ERG(1,4).

The fusion of an ARG to an ETS family member gene is detectable as DNA, RNA, or protein. Initially, the gene fusion is detectable as a chromosomal rearrangement of genomic DNA having a 5’ portion from a transcriptional regulatory region of the ARG and a 3’ portion from the ETS family member gene. Once transcribed, the gene fusion is detectable as a chimeric mRNA having a 3’ portion from the transcriptional regulatory region of the ARG and a 3’ portion from the ETS family member gene. Once translated, the gene fusion is detectable as an amino-terminally truncated ETS family member protein resulting from the fusion of the transcriptional regulatory region of the ARG to the ETS family member gene; a chimeric protein having an amino-terminal portion from the transcriptional regulatory region of the ARG and a carboxy-terminal portion from the ETS family member gene; or, an upregulated, but otherwise indistinguishable, native ETS family member protein. The truncated ETS family member protein and chimeric protein may differ from their respective native proteins in amino acid sequence, post-translational processing and/or secondary, tertiary or quaternary structure. Such differences, if present, can be used to identify the presence of the gene fusion. Specific methods of detection are described in more detail below.

Certain gene fusions are more common than others in prostate cancer. The present invention identifies 50-80% of prostate cancers as having recurrent gene fusions of TMPRSS2 with ERG, ETV1, ETV4, or FLI1. Of these, 50-70% are TMPRSS2:ERG, 50%-60% of which result from the deletion of genetic information between the TMPRSS2 and ERG locus on chromosome 21 (described in more detail below), 5-10% are TMPRSS2-ETV1, 1-2% are TMPRSS2-ETV4, and 1-2% are TMPRSS2-FLI1.

Experiments conducted during the course of development of the present invention indicated that certain fusion genes express fusion transcripts, while others do not express a functional transcript (Tomlins et al., Science, 310: 644-648 (2005); Tomlins et al., Cancer Research 66: 3396-3400 (2006)).

Further experiments conducted during the course of development of the present invention identified significant genomic deletions located between TMPRSS2 and ERG on chromosome 21q22.2-3. Deletions were seen in TMPRSS2: ERG fusion positive PCA samples. The deletions appear in a consensus area but show variability within this area. In previously published work by Paris et al. (Hum. Mol. Genet. 13:303-13 (2004)), CGH analysis detected deletions in the CTD-210307 BAC that is 6 kb centromeric from TMPRSS2.
These deletions were observed in 12.5% (4/32) of clinically localized PCA samples and 33% (9/27) of the metastatic PCA samples. These results support the SNP array data from the current study and suggest that either PCA deletions become more common with progression or that deletions are identified more often in PCA that tend to progress more rapidly. Given the striking intra-tumoral homogeneity of the ATM625:ERG rearrangements, it is more likely that these molecular sub-types are associated with different disease progression characteristics.

One hundred eighteen clinically localized PCA cases with 49.2% harboring rearrangement of ERG were evaluated. Intrinsic deletions were observed in 60.3% of these ATM625:ERG fusion positive cases. Almost all PCA samples with marked over expression of ERG have a rearrangement, and the over expression occurs in about the same number of cases as the rearrangement. Using Oncome, a publicly available compendium of gene expression data, 8 significantly down regulated genes located in the area of the common deletion site were identified (Fig. 16).

The present invention is not limited to a particular mechanism. Indeed, an understanding of the mechanism is not necessary to practice the present invention. Nonetheless, the results suggest that nearly all of all PCAs can be defined by the ATM625:ERG rearrangement. The majority of these tumors demonstrate an intrinsic deletion, which according to the oligonucleotide SNP array genomic analysis is variable in size. However, approximately 30-40% did not demonstrate a deletion and thus might harbor a balanced translocation of ATM625 and ERG. This variability in the extent of the deletion may be associated with disease progression as has been observed with CML. The current study identified significant clinical associations with tumor stage and lymph node status. ATM625:ERG rearranged tumors with deletion also showed a trend towards higher rates of PSA biochemical failure.

Additional experiments conducted during the course of development of the present invention explored the risk of developing metastases or prostate cancer specific death based on the presence of the ATM625:ERG gene fusion in a watchful waiting cohort of early prostate cancers with long term follow-up. The frequency of the ATM625:ERG gene fusion was assessed using 92 cases. The frequency of ATM625:ERG gene fusion in this population-based cohort was 15.2% (14/92), lower than the 50% frequency observed in two hospital-based cohorts. The present invention is not limited to a particular mechanism. Indeed, an understanding of the mechanism is not necessary to practice the present invention. Nonetheless, this difference in ATM625:ERG gene fusion prostate cancers may be due to ethnic and racial genetic differences. These differences may also be explained by the lower percentage of high grade cases in this watchful waiting cohort as compared to the other non-population based studies.

A significant association between ATM625:ERG gene fusion and development of distant metastases and prostate cancer specific death was observed with a cumulative incidence ratio of 3.6 (P=0.004, 95% confidence interval=1.5 to 8.9). These data suggest that ATM625:ERG gene fusion prostate cancers have a more aggressive phenotype. Further experiments indicated that genomic deletions in the ATM625:ERG gene fusion were correlated with advanced and/or metastatic prostate cancer (See e.g., Example 5).

The present invention has also demonstrated that androgen can induce the overexpression of ERG, presumably through ARES, in a ATM625:ERG-positive cell line. The present invention is not limited to a particular mechanism. Indeed, an understanding of the mechanism is not necessary to practice the present invention. Nonetheless, the results suggest that dysregulation of ETS family activity through ARES upstream of ATM625 may drive prostate cancer development.

It is contemplated that the presence, molecular sub-type or amount of gene fusion expression is correlated with the stage, aggressiveness or progression of the disease, or the presence or risk of metastasis. It is further contemplated that similar recurrent gene fusions involving ETS family member genes occur in other epithelial cancers.

II. Antibodies

The gene fusion proteins of the present invention, including fragments, derivatives and analog thereof, may be used as immunogens to produce antibodies having use in the diagnostic, research, and therapeutic methods described below. The antibodies may be polyclonal or monoclonal, chimeric, humanized, single chain or Fab fragments. Various procedures known to those of ordinary skill in the art may be used for the production and labeling of such antibodies and fragments. See, e.g., Burns, ed., Immunological Protocols, 3rd ed., Humana Press (2005); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1988); Kozbor et al., Immunology Today 4: 72 (1983); Köhler and Milstein, Nature 256: 495 (1975). Antibodies or fragments exploiting the differences between the truncated ETS family member protein or chimeric protein and their respective native proteins are particularly preferred.

III. Diagnostic Applications

The present invention provides DNA, RNA and protein based diagnostic methods that either directly or indirectly detect the gene fusions. The present invention also provides compositions and kits for diagnostic purposes.

The diagnostic methods of the present invention may be qualitative or quantitative. Quantitative diagnostic methods may be used, for example, to discriminate between indolent and aggressive cancers via a cutoff or threshold level. Where applicable, qualitative or quantitative diagnostic methods may also include amplification of target, signal or intermedi-ary (e.g., a universal primer).

An initial assay may confirm the presence of a gene fusion but not identify the specific fusion. A secondary assay is then performed to determine the identity of the particular fusion, if desired. The second assay may use a different detection technology than the initial assay.

The gene fusions of the present invention may be detected along with other markers in a multiplex or panel format. Markers are selected for their predictive value alone or in combination with the gene fusions. Exemplary prostate cancer markers include, but are not limited to: AMACR/P504S (U.S. Pat. No. 6,262,245); PC3A (U.S. Pat. No. 7,008,765); PCGEM1 (U.S. Pat. No. 6,828,429); prostateP501S, P503S, P504S, P509S, P510S, prostateP703P, P710P (U.S. Publication No. 20030185830); and, those disclosed in U.S. Pat. Nos. 5,854,206 and 6,034,218, and U.S. Publication No. 20030175736, each of which is herein incorporated by reference in its entirety. Markers for other cancers, diseases, infections, and metabolic conditions are also contemplated for inclusion in a multiplex of panel format.

The diagnostic methods of the present invention may also be modified with reference to data correlating particular gene fusions with the stage, aggressiveness or progression of the disease or the presence or risk of metastasis. Ultimately, the information provided by the methods of the present invention will assist a physician in choosing the best course of treatment for a particular patient.
A. Sample
Any patient sample suspected of containing the gene fusions may be tested according to the methods of the present invention. By way of non-limiting examples, the sample may be tissue (e.g., a prostate biopsy sample or a tissue sample obtained by prostatectomy), blood, urine, semen, prostatic secretions or a fraction thereof (e.g., plasma, serum, urine supernatant, urine cell pellet or prostate cells). A urine sample is preferably collected immediately following an attentive digital rectal examination (DRE), which causes prostate cells from the prostate gland to shed into the urinary tract.

The patient sample typically requires preliminary processing designed to isolate or enrich the sample for the gene fusions or cells that contain the gene fusions. A variety of techniques known to those of ordinary skill in the art may be used for this purpose, including but not limited to: centrifugation; immunocapture; cell lysis; and, nucleic acid target capture (See, e.g., EP Pat. No. 1 409 727, herein incorporated by reference in its entirety).

B. DNA and RNA Detection
The gene fusions of the present invention may be detected as chromosomal rearrangements of genomic DNA or chimeric mRNA using a variety of nucleic acid techniques known to those of ordinary skill in the art, including but not limited to: nucleic acid sequencing; nucleic acid hybridization; and, nucleic acid amplification.

1. Sequencing
Illustrative non-limiting examples of nucleic acid sequencing techniques include, but are not limited to, chain terminator (Sanger) sequencing and dye terminator sequencing. Those of ordinary skill in the art will recognize that because RNA is less stable in the cell and more prone to nuclease attack experimentally RNA is usually reverse transcribed to DNA before sequencing.

Chain terminator sequencing uses sequence-specific termination of a DNA synthesis reaction using modified nucleotide substrates. Extension is initiated at a specific site on the template DNA by using a short radioactive, or other labeled, oligonucleotide primer complementary to the template at that region. The oligonucleotide primer is extended using a DNA polymerase, standard four deoxynucleotide bases, and a low concentration of one chain terminating nucleotide, most commonly a di-deoxyribonucleotide. This reaction is repeated in four separate tubes with each of the bases taking turns as the di-deoxyribonucleotide. Limited incorporation of the chain terminating nucleotide by the DNA polymerase results in a series of related DNA fragments that are terminated only at positions where that particular di-deoxyribonucleotide is used.

For each reaction tube, the fragments are size-separated by electrophoresis in a slab polyacrylamide gel or a capillary tube filled with a viscous polymer. The sequence is determined by reading which lane produces a visualized mark from the labeled primer as you scan from the top of the gel to the bottom.

Dye terminator sequencing alternatively labels the terminators. Complete sequencing can be performed in a single reaction by labeling each of the di-deoxyribonucleotide chain-terminators with a separate fluorescent dye, which fluoresces at a different wavelength.

2. Hybridization
Illustrative non-limiting examples of nucleic acid hybridization techniques include, but are not limited to, in situ hybridization (ISH), microarray, and Southern or Northern blot.

In situ hybridization (ISH) is a type of hybridization that uses a labeled complementary DNA or RNA strand as a probe to localize a specific DNA or RNA sequence in a portion or section of tissue (in situ), or, if the tissue is small enough, the entire tissue (whole mount ISH). DNA ISH can be used to determine the structure of chromosomes. RNA ISH is used to measure and localize mRNAs and other transcripts within tissue sections or whole mounts. Sample cells and tissues are usually treated to fix the target transcripts in place and to increase access of the probe. The probe hybridizes to the target sequence at elevated temperature, and then the excess probe is washed away. The probe that was labeled with either radiolabeled bases is localized and quantitated in the tissue using autoradiography; fluorescence microscopy or immunohistochemistry, respectively. ISH can also use two or more probes, labeled with radioactivity or the other non-radioactive labels, to simultaneously detect two or more transcripts.

2.1 FISH
In some embodiments, fusion sequences are detected using fluorescence in situ hybridization (FISH). The preferred FISH assays for the present invention utilize bacterial artificial chromosomes (BACs). These have been used extensively in the human genome sequencing project (see Nature 409: 953-958 (2001)) and clones containing specific BACs are available through distributors that can be located through many sources, e.g., NCBI. Each BAC clone from the human genome has been given a reference name that unambiguously identifies it. These names can be used to find a corresponding GenBank sequence and to order copies of the clone from a distributor.

In some embodiments, the detection assay is a FISH assay utilizing a probe for ETV1 (e.g., bac RP11-692L4), a set of probes for c-ERG:t-ERG break apart (e.g., bac RP11-24A11 and as a probe for t-ERG RP11-372O17 or RP11-137I13). In other embodiments, the FISH assay is performed by testing for ETV1 deletion or amplification with a set of probes, wherein one probe spans the ETV1 locus (e.g., bac RP11-692L4) and the other probe hybridizes to chromosome 7 (e.g., a probe on the centromere of the chromosome). In still further embodiments, the method is performed by testing for ERG deletion or amplification with a set of probes, one spanning the ERG locus (e.g., bac RP11-476D17) and one reference probe on chromosome 21 (e.g., PR11-32L6; RP11-752M23; RP11-1107I121; RP11-639A7 or RP11-1077M21). In yet other embodiments, the method is performed by testing for TMPRSS2 deletion/amplification with a set of probes, one spanning the TMPRSS2 (e.g., RP11-121A5; RP11-120C17; RR11-814F13; or RR11-535H11) locus and one reference probe on chromosome 21 (e.g., PR11-32L6; RP11-752M23; RP11-1107I121; RP11-639A7 or RP11-1077M21). In some embodiments, the method further comprises a hybridization using a probe selected from the group including, but not limited to RP11-121A5; RP11-120C17; PR11-814F13; and RR11-535H11.

The present invention further provides a method of performing a FISH assay on human prostate cells, human prostate tissue or on the fluid surrounding said human prostate cells or human prostate tissue. In some embodiments, the assay comprises a hybridization step utilizing a probe selected from the group including, but not limited to, RP11-372O17; RP11-137I13; RP11-692L4; RP11-476D17; PR11-32L6; RP11-752M23; RP11-1107I121; RP11-639A7; RP11-1077M21; RP11-121A5; RP11-120C17; PR11-814F13; and RR11-535H11.

Specific BAC clones that can be used in FISH protocols to detect rearrangements relevant to the present invention are as follows:
For testing for an ETV1-TMPRSS2 fusion, one probe spanning the ETV1 and one spanning the TMPRSS2 locus may be used:

BAC for ETV1: RP11-692L4
BAC for TMPRSS2: RP11-121A5, (RP11-120C17, RP11-814F13, RR11-535H11)

Testing ERG translocation with set of probes for c-ERG:
- t-ERG break apart:
  - BAC for c-ERG: RP11-24A11
  - BACs for t-ERG: RP11-372017, RP11-137J13
Testing ETV1 deletion/amplification with set of probes, one spanning the ETV1 locus and one reference probe on chromosome 7:
- BAC for ETV1: RP11-692L4
Testing ERG deletion/amplification with set of probes, one spanning the ERG locus and one reference probe on chromosome 21:
- BAC for ERG: RP11-476D17
BACs for reference probe on chromosome 21:
- Testing TMPRSS2 deletion/amplification with set of probes, one spanning the TMPRSS2 locus and one reference probe on chromosome 21:
  - BACs for TMPRSS2: RP11-121A5, (RP11-120C17, PR11-814F13, RR11-535H11)

Regarding methodology may be obtained from many references including: In situ Hybridization: Medical Applications (eds. G. R. Coulton and J. de Belleroche), Kluwer Academic Publishers, Boston (1992); In situ Hybridization: In Neurobiology: Advances in Methodology (eds. J. H. Eberwine, K. L. Valentin, and J. D. Barchas), Oxford University Press Inc., England (1994); In situ Hybridization: A Practical Approach (ed. D. G. Wilkinson), Oxford University Press Inc., England (1992); Kuo, et al., Am. J. Hum. Genet. 49:112-119 (1991); Klinge, et al., Am. J. Hum. Genet. 51:55-65 (1992); and Ward, et al., Am. J. Hum. Genet. 52:854-865 (1993)). There are also kits that are commercially available and that provide protocols for performing FISH assays (available from e.g., Oncor, Inc., Gaithersburg, Md.). Patents providing guidance on methodology include U.S. Pat. Nos. 5,225,326; 5,545,524; 6,121,489 and 6,573,043. All of these references are hereby incorporated by reference in their entirety and may be used along with similar references in the art and with the information provided in the Examples section herein to establish procedural steps convenient for a particular laboratory.

Table 3 below shows additional BAC clones that find use as FISH probes.

<table>
<thead>
<tr>
<th>Table 3</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>Chromosome</td>
<td>RefSeq</td>
<td>5' BAC</td>
<td>3' BAC</td>
</tr>
<tr>
<td>EBF</td>
<td>11p13</td>
<td>NM_012153</td>
<td>RP5-1135K18</td>
<td>RP5-1002E13</td>
</tr>
<tr>
<td>ELF3</td>
<td>13q14</td>
<td>NM_172737</td>
<td>RP11-8864</td>
<td>RP11-535D9</td>
</tr>
<tr>
<td>ELF2</td>
<td>4q28</td>
<td>NM_201999.1</td>
<td>RP11-226A8</td>
<td>RP11-375P1</td>
</tr>
<tr>
<td>ELF3</td>
<td>1q32</td>
<td>NM_004433</td>
<td>RP11-253B7</td>
<td>RP11-246D15</td>
</tr>
<tr>
<td>ELF4</td>
<td>Xq25</td>
<td>NM_001421</td>
<td>RP5-857H13</td>
<td>RP4-73S9</td>
</tr>
<tr>
<td>ELF5</td>
<td>11p13</td>
<td>NM_001422.2</td>
<td>RP5-1002E13</td>
<td>RP5-1135K18</td>
</tr>
<tr>
<td>ELK1</td>
<td>Xp11</td>
<td>NM_005229</td>
<td>RP1-54B20</td>
<td>RP1-306D1</td>
</tr>
<tr>
<td>ELK3</td>
<td>12q22</td>
<td>NM_005229</td>
<td>RP11-69E3</td>
<td>RP11-510I5</td>
</tr>
<tr>
<td>ELK4</td>
<td>1q32</td>
<td>NM_001973.2</td>
<td>RP11-131E5</td>
<td>RP11-249H15</td>
</tr>
<tr>
<td>ERF</td>
<td>19q3</td>
<td>NM_006494.1</td>
<td>RP11-208G3</td>
<td>RP11-317E13</td>
</tr>
<tr>
<td>ERG</td>
<td>21q22</td>
<td>NM_004440.3</td>
<td>RP11-137J13</td>
<td>RP11-24A11</td>
</tr>
<tr>
<td>ETS1</td>
<td>11q24</td>
<td>NM_005238.2</td>
<td>RP11-284C5</td>
<td>RP11-11n22</td>
</tr>
<tr>
<td>ETS2</td>
<td>21q22</td>
<td>NM_005239.4</td>
<td>RP11-24A11</td>
<td>RP11-375G13</td>
</tr>
<tr>
<td>ETV1</td>
<td>7p21</td>
<td>NM_004956.3</td>
<td>RP11-140I13</td>
<td>RP11-34E22</td>
</tr>
<tr>
<td>ETV2</td>
<td>15q13</td>
<td>NM_014209.1</td>
<td>RP11-32I17</td>
<td>RP11-92Q4</td>
</tr>
<tr>
<td>ETV3</td>
<td>1q23</td>
<td>NM_005240.1</td>
<td>RP11-91G5</td>
<td>RP11-103B13</td>
</tr>
<tr>
<td>ETV4</td>
<td>17q21</td>
<td>NM_001986.1</td>
<td>RP11-43H6</td>
<td>RP11-100E5</td>
</tr>
<tr>
<td>ETV5</td>
<td>3q27</td>
<td>NM_004454.1</td>
<td>RP11-379C23</td>
<td>RP11-144N13</td>
</tr>
<tr>
<td>ETV6</td>
<td>12p13</td>
<td>NM_004987.3</td>
<td>RP11-90N7</td>
<td>RP11-99L1</td>
</tr>
<tr>
<td>ETV7</td>
<td>6p21</td>
<td>NM_016135.2</td>
<td>RP3-43H14</td>
<td>RP11-179N16</td>
</tr>
<tr>
<td>FEB</td>
<td>2q35</td>
<td>NM_017521.2</td>
<td>RP11-310D14</td>
<td>RP11-29D2</td>
</tr>
<tr>
<td>FLI1</td>
<td>11q24</td>
<td>NM_002017.2</td>
<td>RP11-11n22</td>
<td>RP11-75P14</td>
</tr>
<tr>
<td>FLJ16478</td>
<td>1q23</td>
<td>NM_00100414</td>
<td>RP11-91G5</td>
<td>RP11-103B13</td>
</tr>
<tr>
<td>SPDEF</td>
<td>6p21</td>
<td>NM_012391.1</td>
<td>RP11-79D3</td>
<td>RP11-119G22</td>
</tr>
<tr>
<td>SPI1</td>
<td>11p11</td>
<td>NM_016135.2</td>
<td>RP11-56E13</td>
<td>RP11-29D2</td>
</tr>
<tr>
<td>SPI1</td>
<td>11q13</td>
<td>NM_001212.2</td>
<td>RP11-501H16</td>
<td>RP11-69P1</td>
</tr>
<tr>
<td>SPIC</td>
<td>12q23</td>
<td>NM_152323.1</td>
<td>RP11-426H24</td>
<td>RP11-93S1C1</td>
</tr>
<tr>
<td>TMPRSS2</td>
<td>21q22</td>
<td>NM_005656.2</td>
<td>RP11-33C4</td>
<td>RP11-120C17</td>
</tr>
</tbody>
</table>


The most preferred probes for detecting a deletion mutation resulting in a fusion between TMPRSS2 and ERG are RP11-24A11 and RP11-137J13. These probes, or those described above, are labeled with appropriate fluorescent or other markers and then used in hybridizations. The Examples section provided herein sets forth one particular protocol that is effective for measuring deletions but one of skill in the art will recognize that many variations of this assay can be used equally well. Specific protocols are well known in the art and can be readily adapted for the present invention. Guidance

2.2 Microarrays

Different kinds of biological assays are called microarrays including, but not limited to: DNA microarrays (e.g., cDNA microarrays and oligonucleotide microarrays); protein microarrays; tissue microarrays; transfection or cell microarrays; chemical compound microarrays; and, antibody microarrays. A DNA microarray, commonly known as gene chip, DNA chip, or biochip, is a collection of microscopic DNA spots attached to a solid surface (e.g., glass, plastic or silicon chip) forming an array for the purpose of expression profiling or monitoring expression levels for thousands of genes simultaneously. The affixed DNA segments are known as probes, thousands of which can be used in a single DNA
microarray. Microarrays can be used to identify disease genes by comparing gene expression in disease and normal cells. Microarrays can be fabricated using a variety of technologies, including but not limited to: printing with fine-pointed pins onto glass slides; photolithography using pre-made masks; photolithography using dynamic micromirror devices; ink jet printing; or, electrochemistry on microelectrode arrays.

Southern and Northern blotting is used to detect specific DNA or RNA sequences, respectively. DNA or RNA extracted from a sample is fragmented, electrophoretically separated on a matrix gel, and transferred to a membrane filter. The filter bound DNA or RNA is subject to hybridization with a labeled probe complementary to the sequence of interest. Hybridized probe bound to the filter is detected. A variant of the procedure is the reverse Northern blot, in which the substrate nucleic acid that is affixed to the membrane is a collection of isolated DNA fragments and the probe is RNA extracted from a tissue and labeled.

3. Amplification

Chromosomal rearrangements of genomic DNA and chimeric mRNA may be amplified prior to or simultaneous with detection. Illustrative non-limiting examples of nucleic acid amplification techniques include, but are not limited to, polymerase chain reaction (PCR), reverse transcription polymerase chain reaction (RT-PCR), transcription-mediated amplification (TMA), ligase chain reaction (LCR), strand displacement amplification (SDA), and nucleic acid sequence based amplification (NASBA). Those of ordinary skill in the art will recognize that certain amplification techniques (e.g., PCR) require that RNA be reversed transcribed to DNA prior to amplification (e.g., RT-PCR), whereas other amplification techniques directly amplify RNA (e.g., TMA and NASBA).

The polymerase chain reaction (U.S. Pat. Nos. 4,683,195, 4,683,202, 4,800,159 and 4,965,198; each of which is herein incorporated by reference in its entirety), commonly referred to as PCR, uses multiple cycles of denaturation, annealing of primer pairs to opposite strands, and primer extension to exponentially increase copy numbers of a target nucleic acid sequence. In a variation called RT-PCR, reverse transcriptase (RT) is used to make a complementary DNA (cDNA) from mRNA, and the cDNA is then amplified by PCR to produce multiple copies of DNA. For other various permutations of PCR see, e.g., U.S. Pat. Nos. 4,683,195, 4,683,202 and 4,800,159; Mullis et al., Meth. Enzymol. 155: 335 (1987); and Murakawa et al., DNA 7: 287 (1988), each of which is herein incorporated by reference in its entirety.

Transcription mediated amplification (U.S. Pat. Nos. 5,480,784 and 5,399,491, each of which is herein incorporated by reference in its entirety), commonly referred to as TMA, synthesizes multiple copies of a target nucleic acid sequence autokatally and under conditions of substantially constant temperature, ionic strength, and pH in which multiple RNA copies of the target sequence autokatally generate additional copies. See, e.g., U.S. Pat. Nos. 5,399,491 and 5,824,518, each of which is herein incorporated by reference in its entirety. In a variation described in U.S. Publ. No. 20060046265 (herein incorporated by reference in its entirety), TMA optionally incorporates the use of blocking moieties, terminating moieties, and other modifying moieties to improve TMA process sensitivity and accuracy.

The ligase chain reaction (Weiss, R., Science 254: 1292 (1991), herein incorporated by reference in its entirety), commonly referred to as LCR, uses two sets of complementary DNA oligonucleotides that hybridize to adjacent regions of the target nucleic acid. The DNA oligonucleotides are covalently linked by a DNA ligase in repeated cycles of thermal denaturation, hybridization and ligation to produce a detectable double-stranded ligated oligonucleotide product.

Strand displacement amplification (Walker, G. et al., Proc. Natl. Acad. Sci. USA 89: 392-396 (1992); U.S. Pat. Nos. 5,270,184 and 5,455,166, each of which is herein incorporated by reference in its entirety), commonly referred to as SDA, uses cycles of annealing of primer sequences to opposite strands of a target sequence, primer extension in the presence of a dNTPs to produce a duplex hemiphiphosphothioate primer extension product, endonuclease-mediated nicking of a hemimodified restriction endonuclease recognition site, and polymerase-mediated primer extension from the 3' end of the nick to displace an existing strand and produce a strand for the next round of primer annealing, nicking and strand displacement, resulting in geometric amplification of product. Thermodhaphic SDA (tSDA) uses thermophilic endonucleases and polymerases at higher temperatures in essentially the same method (EP Pat. No. 0 684 315).

Other amplification methods include, for example: nucleic acid sequence based amplification (U.S. Pat. No. 5,130,238; herein incorporated by reference in its entirety); commonly referred to as NASBA; one that uses an RNA replicase to amplify the probe molecule itself (Lizardi et al., BioTechnol. 6: 1197 (1988), herein incorporated by reference in its entirety), commonly referred to as Qβ replicase; a transcription based amplification method (Kwok et al., Proc. Natl. Acad. Sci. USA 86:1173 (1989)); and, self-sustained sequence replication (Guatelli et al., Proc. Natl. Acad. Sci. USA 87: 1874 (1990), each of which is herein incorporated by reference in its entirety). For further discussion of known amplification methods see Persing, David H., “In Vitro Nucleic Acid Amplification Techniques” in Diagnostic Medical Microbiology: Principles and Applications (Persing et al., Eds.), pp. 51-87 (American Society for Microbiology, Washington, D.C. (1993)).

4. Detection Methods

Non-amplified or amplified gene fusion nucleic acids can be detected by any conventional means. For example, the gene fusions can be detected by hybridization with a detectably labeled probe and measurement of the resulting hybrids. Illustrative non-limiting examples of detection methods are described below.

One illustrative detection method, the Hybridization Protection Assay (HPA) involves hybridizing a chemiluminescent oligonucleotide probe (e.g., an acidinium ester-labeled (AE) probe) to the target sequence, selectively hydrolyzing the chemiluminescent label present on unhybridized probe, and measuring the chemiluminescence produced from the remaining probe in a luminometer. See, e.g., U.S. Pat. No. 5,203,174 and Norman C. Nelson et al., Nonisotopic Probing, Blotting, and Sequencing, ch. 17 (Larry J. Krizka ed., 2d ed. 1995, each of which is herein incorporated by reference in its entirety).

Another illustrative detection method provides for quantitative evaluation of the amplification process in real-time. Evaluation of an amplification process in “real-time” involves determining the amount of amplicon in the reaction mixture either continuously or periodically during the amplification reaction, and using the determined values to calculate the amount of target sequence initially present in the sample. A variety of methods for determining the amount of initial target sequence present in a sample based on real-time amplification are well known in the art. These include methods disclosed in U.S. Pat. Nos. 6,303,305 and 6,541,205, each of which is herein incorporated by reference in its entirety. Another method for determining the quantity of target sequence initially present in a sample, but which is not based on a real-
time amplification, is disclosed in U.S. Pat. No. 5,710,029, herein incorporated by reference in its entirety.

Amplification products may be detected in real-time through the use of various self-hybridizing probes, most of which have a stem-loop structure. Such self-hybridizing probes are labeled so that they emit differently detectable signals, depending on whether the probes are in a self-hybridized state or an altered state through hybridization to a target sequence. By way of non-limiting example, “molecular torches” are a type of self-hybridizing probe that includes distinct regions of self-complementarity (referred to as “the target binding domain” and “the target closing domain”) which are connected by a joining region (e.g., non-nucleotide linker) and which hybridize to each other under predetermined hybridization assay conditions. In a preferred embodiment, molecular torches contain single-stranded base regions in the target binding domain that range from 1 to about 20 bases in length and are accessible for hybridization to a target sequence present in an amplification reaction under strand displacement conditions. Under strand displacement conditions, hybridization of the two complementary regions, which may be fully or partially complementary, of the molecular torch is favored, except in the presence of the target sequence, which will bind to the single-stranded region present in the target binding domain and displace all or a portion of the target closing domain. The target binding domain and the target closing domain of a molecular torch include a detectable label or a pair of interacting labels (e.g., luminescent quencher) positioned so that a different signal is produced when the molecular torch is self-hybridized than when the molecular torch is hybridized to the target sequence, thereby permitting detection of probe/target duplexes in a test sample in the presence of unhybridized molecular torches. Molecular torches and a variety of types of interacting label pairs are disclosed in U.S. Pat. No. 6,534,274, herein incorporated by reference in its entirety. Another example of a detection probe having self-complementarity is a “molecular beacon.” Molecular beacons include nucleic acid molecules having a target complementary sequence, an unlabeled pair (or nucleic acid arms) holding the probe in a closed conformation in the absence of a target sequence present in an amplification reaction, and a label pair that interacts when the probe is in a closed conformation. Hybridization of the target sequence and the target complementary sequence separates the members of the affinity pair, thereby shifting the probe to an open conformation. The shift to the open conformation is detectable due to reduced interaction of the label pair, which may be, for example, a fluorophore and a quencher (e.g., DABCYL, and EDANS). Molecular beacons are disclosed in U.S. Pat. Nos. 5,925,517 and 6,150,097, herein incorporated by reference in its entirety. Other self-hybridizing probes are well known to those of ordinary skill in the art. By way of non-limiting example, probe binding pairs having interacting labels, such as those disclosed in U.S. Pat. No. 5,928,862 (herein incorporated by reference in its entirety) might be adapted for use in the present invention. Probe systems used to detect single nucleotide polymorphisms (SNPs) might also be utilized in the present invention. Additional detection systems include “molecular switches,” as disclosed in U.S. Pat. No. 20050042638, herein incorporated by reference in its entirety. Other probes, such as those comprising intercalating dyes and/or fluorochromes, are also useful for detection of amplification products in the present invention. See, e.g., U.S. Pat. No. 5,814,447 (herein incorporated by reference in its entirety).

C. Protein Detection

The gene fusions of the present invention may be detected as truncated ETS family member proteins or chimeric proteins using a variety of protein techniques known to those of ordinary skill in the art, including but not limited to: protein sequencing; and, immunosassays.

1. Sequencing

Illustrative non-limiting examples of protein sequencing techniques include, but are not limited to, mass spectrometry and Edman degradation.

Mass spectrometry can, in principle, sequence any size protein but becomes computationally more difficult as size increases. A protein is digested by an endoprotease, and the resulting solution is passed through a high pressure liquid chromatography column. At the end of this column, the solution is sprayed out of a narrow nozzle charged to a high positive potential into the mass spectrometer. The charge on the droplets causes them to fragment until only single ions remain. The peptides are then fragmented and the mass-charge ratios of the fragments measured. The mass spectrum is analyzed by computer and often compared against a database of previously sequenced proteins in order to determine the sequences of the fragments. The process is then repeated with a different digestion enzyme, and the overlaps in sequences are used to construct a sequence for the protein.

In the Edman degradation reaction, the peptide to be sequenced is adsorbed onto a solid surface (e.g., a glass fiber coated with polybrene). The Edman reagent, phenylisothiocyanate (PTC), is added to the adsorbed peptide, together with a mildly basic buffer solution of 12% trimethylamine, and reacts with the amine group of the N-terminal amino acid. The terminal amino acid derivative can then be selectively extracted by the addition of anhydrous acid. The derivative isomerizes to give a substituted phenylthiohydantoin, which can be washed off and identified by chromatography, and the cycle can be repeated. The efficiency of each step is about 98%, which allows about 50 amino acids to be reliably determined.

2. Immunosassays

Illustrative non-limiting examples of immunosassays include, but are not limited to: immunoprecipitation, Western blot; ELISA; immunohistochemistry; immunocytochemistry; flow cytometry; and, immuno-PCR. Polyclonal or monoclonal antibodies detectably labeled using various techniques known to those of ordinary skill in the art (e.g., colorimetric, fluorescent, chemiluminescent or radioactive) are suitable for use in the immunosassays.

Immunoprecipitation is the technique of precipitating an antigen out of solution using an antibody specific to that antigen. The process can be used to identify protein complexes present in cell extracts by targeting a protein believed to be in the complex. The complexes are brought out of solution by insoluble antibody-binding proteins isolated initially from bacteria, such as Protein A and Protein G. The antibodies can also be coupled to sepharose beads that can easily be isolated out of solution. After washing, the precipitate can be analyzed using mass spectrometry, Western blotting, or any number of other methods for identifying constituents in the complex.

A Western blot, or immunoblot, is a method to detect protein in a given sample of tissue homogenate or extract. It uses gel electrophoresis to separate denatured proteins by mass. The proteins are then transferred out of the gel and onto a membrane, typically polyvinylidene fluoride or nitrocellulose, where they are probed using antibodies specific to the protein
of interest. As a result, researchers can examine the amount of protein in a given sample and compare levels between several groups.

An ELISA, short for Enzyme-Linked ImmunoSorbent Assay, is a biochemical technique to detect the presence of an antibody or an antigen in a sample. It utilizes a minimum of two antibodies, one of which is specific to the antigen and the other of which is coupled to an enzyme. The second antibody will cause a chromogenic or fluorogenic substrate to produce a signal. Variations of ELISA include sandwich ELISA, competitive ELISA, and ELISPOT. Because the ELISA can be performed to evaluate either the presence of antigen or the presence of antibody in a sample, it is a useful tool both for determining serum antibody concentrations and also for detecting the presence of antigen.

Immunohistochemistry and immunocytochemistry refer to the process of localizing proteins in a tissue section or cell, respectively, via the principle of antigens in tissue or cells binding to their respective antibodies. Visualization is enabled by tagging the antibody with color producing or fluorescent tags. Typical examples of color tags include, but are not limited to, horseradish peroxidase and alkaline phosphatase. Typical examples of fluorophore tags include, but are not limited to, fluorescein isothiocyanate (FITC) or phycoerythrin (PE).

Flow cytometry is a technique for counting, examining and sorting microscopic particles suspended in a stream of fluid. It allows simultaneous multiparametric analysis of the physical and/or chemical characteristics of single cells flowing through an optical/electronic detection apparatus. A beam of light (e.g., a laser) of a single frequency or color is directed onto a hydrodynamically focused stream of fluid. A number of detectors are aimed at the point where the stream passes through the light beam; one in line with the light beam (Forward Scatter or FSC) and several perpendicular to it (Side Scatter or SSC) and one or more fluorescent detectors. Each suspended particle passing through the beam scatters the light in some way, and fluorescent chemicals in the particle may be excited into emitting light at a lower frequency than the light source. The combination of scattered and fluorescent light is picked up by the detectors, and by analyzing fluctuations in brightness at each detector, one for each fluorescent emission peak, it is possible to deduce various facts about the physical and chemical structure of each individual particle. FSC correlates with the cell volume and SSC correlates with the density or inner complexity of the particle (e.g., shape of the nucleus, the amount and type of cytoplasmic granules or the membrane roughness).

Immunopolymerase chain reaction (IPCR) utilizes nucleic acid amplification techniques to increase signal generation in antibody-based immunosays. Because no protein equivalent of PCR exists, that is, proteins cannot be replicated in the same manner that nucleic acid is replicated during PCR, the only way to increase detection sensitivity is by signal amplification. The target proteins are bound to antibodies which are directly or indirectly conjugated to oligonucleotides. Unbound antibodies are washed away and the remaining bound antibodies have their oligonucleotides amplified. Protein detection occurs via detection of amplified oligonucleotides using standard nucleic acid detection methods, including real-time methods.

D. Data Analysis
In some embodiments, a computer-based analysis program is used to translate the raw data generated by the detection assay (e.g., the presence, absence, or amount of a given marker or markers) into data of predictive value for a clinician. The clinician can access the predictive data using any suitable means. Thus, in some preferred embodiments, the present invention provides the further benefit that the clinician, who is not likely to be trained in genetics or molecular biology, need not understand the raw data. The data is presented directly to the clinician in its most useful form. The clinician is then able to immediately utilize the information in order to optimize the care of the subject.

The present invention contemplates any method capable of receiving, processing, and transmitting the information to and from laboratories conducting the assays, information provides, medical personal, and subjects. For example, in some embodiments of the present invention, a sample (e.g., a biopsy or a serum or urine sample) is obtained from a subject and submitted to a profiling service (e.g., clinical lab at a medical facility, genomic profiling business, etc.), located in any part of the world (e.g., in a country different than the country where the subject resides or where the information is ultimately used) to generate raw data. Where the sample comprises a tissue or other biological sample, the subject may visit a medical center to have the sample obtained and sent to the profiling center, or subjects may collect the sample themselves (e.g., a urine sample) and directly send it to a profiling center. Where the sample comprises previously determined biological information, the information may be directly sent to the profiling service by the subject (e.g., an information card containing the information may be scanned by a computer and the data transmitted to a computer of the profiling center using an electronic communication systems). Once received by the profiling service, the sample is processed and a profile is produced (i.e., expression data), specific for the diagnostic or prognostic information desired for the subject.

The profile data is then prepared in a format suitable for interpretation by a treating clinician. For example, rather than providing raw expression data, the prepared format may represent a diagnosis or risk assessment (e.g., likelihood of cancer being present) for the subject, along with recommendations for particular treatment options. The data may be displayed to the clinician by any suitable method. For example, in some embodiments, the profiling service generates a report that can be printed for the clinician (e.g., at the point of care) or displayed to the clinician on a computer monitor.

In some embodiments, the information is first analyzed at the point of care or at a regional facility. The raw data is then sent to a central processing facility for further analysis and/or to convert the raw data to information useful for a clinician or patient. The central processing facility provides the advantage of privacy (all data is stored in a central facility with uniform security protocols), speed, and uniformity of data analysis. The central processing facility can then control the fate of the data following treatment of the subject. For example, using an electronic communication system, the central facility can provide data to the clinician, the subject, or researchers.

In some embodiments, the subject is able to directly access the data using the electronic communication system. The subject may chose further intervention or counseling based on the results. In some embodiments, the data is used for research use. For example, the data may be used to further optimize the inclusion or elimination of markers as useful indicators of a particular condition or stage of disease.

E. In Vivo Imaging
The gene fusions of the present invention may also be detected using in vivo imaging techniques, including but not limited to: radionuclide imaging; positron emission tomography (PET); computerized axial tomography. X-ray or magnetic resonance imaging method, fluorescence detection, and
chemiluminescent detection. In some embodiments, in vivo imaging techniques are used to visualize the presence of or expression of cancer markers in an animal (e.g., a human or non-human mammal). For example, in some embodiments, cancer marker mRNA or protein is labeled using a labeled antibody specific for the cancer marker. A specifically bound and labeled antibody can be detected in an individual using an in vivo imaging method, including, but not limited to, radiouclide imaging, positron emission tomography, computerized axial tomography, X-ray or magnetic resonance imaging method, fluorescence detection, and chemiluminescent detection. Methods for generating antibodies to the cancer markers of the present invention are described below.

The in vivo imaging methods of the present invention are useful in the diagnosis of cancers that express the cancer markers of the present invention (e.g., prostate cancer). In vivo imaging is used to visualize the presence of a marker indicative of the cancer. Such techniques allow for diagnosis without the use of an unpleasant biopsy. The in vivo imaging methods of the present invention are also useful for providing prognoses to cancer patients. For example, the presence of a marker indicative of cancers likely to metastasize can be detected. The in vivo imaging methods of the present invention can further be used to detect metastatic cancers in other parts of the body.

In some embodiments, reagents (e.g., antibodies) specific for the cancer markers of the present invention are fluorescently labeled. The labeled antibodies are introduced into a subject (e.g., orally or parenterally). Fluorescently labeled antibodies are detected using any suitable method (e.g., using the apparatus described in U.S. Pat. No. 6,198,107, herein incorporated by reference).

In other embodiments, antibodies are radioactively labeled. The use of antibodies for in vivo diagnosis is well known in the art. Smerdon et al., (Nucl. Med. Biol. 17:247-254 [1990]) have described an optimized antibody-chelator for the radioimmunoscintigraphic imaging of tumors using Indium-111 as the label. Griffin et al., (J Clin Ono 9:631-640 [1991]) have described the use of this agent in detecting tumors in patients suspected of having recurrent colorectal cancer. The use of similar agents with paramagnetic ions as labels for magnetic resonance imaging is known in the art (Lawther, Magnetic Resonance in Medicine 22:339-342 [1991]). The label used will depend on the imaging modality chosen. Radioactive labels such as Indium-111, Technetium-99m, or Iodine-131 can be used for planar scans or single photon emission computed tomography (SPECT). Positron emitting labels such as Fluorine-18 can also be used for positron emission tomography (PET). For MRI, paramagnetic ions such as Gadolinium (III) or Manganese (II) can be used.

Radioactive metals with half-lives ranging from 1 hour to 3.5 days are available for conjugation to antibodies, such as scandium-47 (3.5 days) gallium-67 (2.8 days), gallium-68 (68 minutes), technetium-99m (6 hours), and indium-111 (3.2 days), of which gallium-67, technetium-99m, and indium-111 are preferably for gamma camera imaging. Gallium-68 is preferable for positron emission tomography.

A useful method of labeling antibodies with such radioisotopes is by means of a bifunctional chelating agent, such as diethylenetriaminepentaacetic acid (DTPA), as described, for example, by Khaw et al. (Science 209:295 [1980]) for In-111 and Tc-99m, and by Scheinberg et al. (Science 215:1511 [1982]). Other chelating agents may also be used, but the 1-(p-carboxymethoxybenzyl)EDTA and the carboxycarbonic anhydride of DTPA are advantageous because their use permits conjugation without affecting the antibody’s immunoreactivity substantially.

Another method for coupling DPTA to proteins is by use of the cyclic anhydride of DTPA, as described by Hinowitch et al. (Int. J. Appl. Radiat. Isot. 33:327 [1982]) for labeling of albumin with In-111, but which can be adapted for labeling of antibodies. A suitable method of labeling antibodies with Tc-99m which does not use chelation with DPTA is the pre-taining method of Crockford et al., (U.S. Pat. No. 4,323,546, herein incorporated by reference).


In the case of the radiometals conjugated to the specific antibody, it is likewise desirable to introduce as high a proportion of the radiolabel as possible into the antibody molecule without destroying its immunospecificity. A further improvement may be achieved by effecting radiolabeling in the presence of the specific cancer marker of the present invention, to insure that the antigen binding site on the antibody will be protected. The antigen is separated after labeling.

In still further embodiments, in vivo biophotonic imaging (Xenogen, Almeda, Calif.) is utilized for in vivo imaging. This real-time in vivo imaging utilizes luciferase. The luciferase gene is incorporated into cells, microorganisms, and animals (e.g., as a fusion protein with a cancer marker of the present invention). When active, it leads to a reaction that emits light. A CCD camera and software is used to capture the image and analyze it.

F. Compositions & Kits
Compositions for use in the diagnostic methods of the present invention include, but are not limited to, probes, amplification oligonucleotides, and antibodies. Particularly preferred compositions detect a product only when an ARG fuses to ETS family member gene. These compositions include: a single labeled probe comprising a sequence that hybridizes to the junction at which a 5' portion from a transcriptional regulatory region of an ARG fuses to a 3' portion from an ETS family member gene (i.e., spans the gene fusion junction); a pair of amplification oligonucleotides wherein the first amplification oligonucleotide comprises a sequence that hybridizes to a transcriptional regulatory region of an ARG and the second amplification oligonucleotide comprises a sequence that hybridizes to an ETS family member gene; an antibody to an amino-terminally truncated ETS family member protein resulting from a fusion of a transcriptional regulatory region of an ARG to an ETS family member gene; or, an antibody to a chimeric protein having an amino-terminal portion from a transcriptional regulatory region of an ARG and a carboxy-terminal portion from an ETS family member gene. Other useful compositions, however, include: a pair of labeled probes wherein the first labeled probe comprises a sequence that hybridizes to a transcriptional regulatory region of an ARG and the second labeled probe comprises a sequence that hybridizes to an ETS family member gene.

Any of these compositions, alone or in combination with other compositions of the present invention, may be provided in the form of a kit. For example, the single labeled probe and pair of amplification oligonucleotides may be provided in a kit for the amplification and detection of gene fusions of the present invention. Kits may further comprise appropriate controls and/or detection reagents.
The probe and antibody compositions of the present invention may also be provided in the form of an array.

IV. Prognostic Applications

Experiments conducted during the course of development of the present invention demonstrated a close correlation between gene fusions of the present invention and the prognosis of patients with prostate cancer (see e.g., Example 5 below). Especially in cases where a fusion results from a deletion of the genomic DNA lying between TMPRSS2 and ERG, it has been found that cancer cells assume a more aggressive phenotype. Thus, in some embodiments, assays that are capable of detecting gene fusions between TMPRSS2 and ERG in which there has been a deletion of intervening DNA are used to provide prognoses and help physicians decide on appropriate therapeutic strategy. For example, in some embodiments, patients with tumors having this particular rearrangement are treated more intensively since their prognosis is significantly worse than patients that lack the rearrangement.

Any assay may be used to determine whether cells are present having a rearrangement of the type discussed above (e.g., those described above).

Although the present invention will most preferably be used in connection with obtaining a prognosis for prostate cancer patients, other epithelial cell tumors may also be examined and the assays and probes described herein may be used in determining whether cancerous cells from these tumors have rearrangements that are likely to make them particularly aggressive, i.e., likely to be invasive and metastatic. Examples of tumors that may be characterized using this procedure include tumors of the breast, lung, colon, ovary, uterus, esophagus, stomach, liver, kidney, brain, skin and muscle. The assays will also be of value to researchers studying these cancers in cell lines and animal models.

Further experiments conducted during the course of development of the present invention demonstrated that chromosomal deletions can be detected by assaying samples to determine whether there is a loss of expression of one or more genes located in the deleted region. For example, approximately 2.8 megabases of genomic DNA is typically deleted in forming a fusion between TMPRSS2 and ERG and at least four genes lying in this area are lost when this occurs. These are the ETS2 gene, the WRB gene, the PTP4 gene and the MX1 gene. A decrease in one or more of these in cancerous prostate cells suggests a poor prognosis.

Accordingly, in some embodiments, the present invention provides a method of assaying epithelial cells for the deletion of chromosomal DNA indicative of a cancer-associated rearrangement, comprising performing a FISH assay using at least a first and a second probe, wherein the first probe is at least 15 nucleotides in length (e.g., at least 15, 20, 35, etc.), is bound to a first fluorescent label; and hybridizes under stringent conditions to a first sequence in the human genome wherein the first sequence includes at least a portion of either an androgen responsive gene (e.g., the TMPRSS2 gene) or a ETS family gene (e.g., the ERG gene, the ETV1 gene, or the ETV4 gene); and the second probe: is at least 15 nucleotides in length; is bound to a second fluorescent label that is different from the first fluorescent label; and hybridizes under stringent conditions to a second sequence in the human genome that is different from the first sequence and which includes at least a portion of an androgen responsive gene (e.g., the TMPRSS2 gene) or a ETS family gene (e.g., the ERG gene, the ETV1 gene, or the ETV4 gene).

In further embodiments, the present invention provides a method for assaying epithelial cells (e.g., prostate cells) for a deletion of genomic DNA indicative of a cancer-associated rearrangement, comprising: obtaining a test sample of epithelial cells; assaying the sample of epithelial cells to determine the level of expression of one or more genes selected from the group including, but not limited to, ETS2; WRB; PTP4; and MX1; comparing the expression level determined in step b) with the level in a control sample; and concluding that a deletion has occurred if the level of expression determined for the gene in the test sample is lower than that for a control sample.

V. Drug Screening Applications

In some embodiments, the present invention provides drug screening assays (e.g., to screen for anticancer drugs). The screening methods of the present invention utilize cancer markers identified using the methods of the present invention (e.g., including but not limited to, ERG, ET1, ETV4, and FLI1 gene fusions with TMPRSS2). For example, in some embodiments, the present invention provides methods of screening for compounds that alter (e.g., decrease) the expression of cancer marker genes. The compounds or agents may interfere with transcription, by interacting, for example, with the promoter region. The compounds or agents may interfere with mRNA produced from the fusion (e.g., by RNA interference, antisense technologies, etc.). The compounds or agents may interfere with pathways that are upstream or downstream of the biological activity of the fusion. In some embodiments, candidate compounds are antisense or interfering RNA agents (e.g., oligonucleotides) directed against cancer markers. In other embodiments, candidate compounds are antibodies or small molecules that specifically bind to a cancer marker regulator or expression products of the present invention and inhibit its biological function.

In one screening method, candidate compounds are evaluated for their ability to alter cancer marker expression by contacting a compound with a cell expressing a cancer marker and then assaying for the effect of the candidate compounds on expression. In some embodiments, the effect of candidate compounds on expression of a cancer marker gene is assayed by detecting the level of cancer marker mRNA expressed by the cell. mRNA expression can be detected by any suitable method. In other embodiments, the effect of candidate compounds on expression of cancer marker genes is assayed by measuring the level of polypeptide encoded by the cancer markers. The level of polypeptide expressed can be measured using any suitable method, including but not limited to, those disclosed herein.

Specifically, the present invention provides screening methods for identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to cancer markers of the present invention, have an inhibitory (or stimulatory) effect on, for example, cancer marker expression or cancer marker activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a cancer marker substrate. Compounds thus identified can be used to modulate the activity of target gene products (e.g., cancer marker genes) either directly or indirectly in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions. Compounds that inhibit the activity or expression of cancer markers are useful in the treatment of proliferative disorders, e.g., cancer, particularly prostate cancer.

In one embodiment, the invention provides assays for screening candidate or test compounds that are substrates of a cancer marker protein or peptidopeptide or a biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds that
31 bind to or modulate the activity of a cancer marker protein or polypeptide or a biologically active portion thereof.

The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone, which are resistant to enzymatic degradation but which nevertheless remain bioactive; see, e.g., Zuckermann et al., J. Med. Chem. 37: 2678-85 [1994]); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are preferred for use with peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) Anticancer Drug Des. 12:145).


In one embodiment, an assay is a cell-based assay in which a cell that expresses a cancer marker mRNA or protein or a biologically active portion thereof is contacted with a test compound, and the ability of the test compound to modulate the activity of a cancer marker is determined. Determining the ability of the test compound to modulate cancer marker activity can be accomplished by monitoring, for example, changes in enzyme activity, transcription or mRNA, or the like.

The ability of the test compound to modulate cancer marker binding to a compound, e.g., a cancer marker substrate or modulator, can also be evaluated. This can be accomplished, for example, by coupling the compound, e.g., the substrate, with a radioligand or enzymatic label such that binding of the compound, e.g., the substrate, to a cancer marker can be determined by detecting the labeled compound, e.g., substrate, in a complex.

Alternatively, the cancer marker is coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate cancer marker binding to the cancer marker substrate in a complex. For example, compounds (e.g., substrates) can be labeled with $^{131}$I, $^{35}$S, 3H, or $^{3}$H, either directly or indirectly, and the radioisotope detected by direct counting of radioactivity or by scintillation counting.

Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

The ability of a compound (e.g., a cancer marker substrate) to interact with a cancer marker with or without the labeling of any of the interactants can be evaluated. For example, a microfluidic reactor can be used to detect the interaction of a compound with a cancer marker without the labeling of either the compound or the cancer marker (McConnell et al., Science 257:1906-1912 [1992]). As used herein, a "microfluidic reactor" (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and cancer markers.

In yet another embodiment, a cell-free assay is provided in which a cancer marker protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the cancer marker protein, mRNA, or biologically active portion thereof is evaluated. Preferred biologically active portions of the cancer marker proteins or mRNA to be used in assays of the present invention include fragments that participate in interactions with substrates or other proteins, e.g., fragments with high surface probability scores.

Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected.

The interaction between two molecules can also be detected, e.g., using fluorescence energy transfer (FRET) (see, for example, Lakowicz et al., U.S. Pat. No. 5,631,169; Stavrianopoulos et al., U.S. Pat. No. 4,968,103; each of which is herein incorporated by reference). A fluorophore label is selected such that a first donor molecule’s emitted fluorescent energy will be absorbed by a fluorescent label on a second, ‘acceptor’ molecule, which in turn is able to fluoresce due to the absorbed energy.

Alternatively, the ‘donor’ protein molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light such that the ‘acceptor’ molecule label may be differentiated from that of the ‘donor’. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the ‘acceptor’ molecule label should be maximal. A FRET binding event can be conveniently measured through standard fluorometric detection means well known in the art (e.g., using a fluorometer).

In another embodiment, determining the ability of the cancer marker protein or mRNA to bind to a target molecule can be accomplished using real-time Molecular Interaction Analysis (MIA) (see, e.g., Sjolander and Urbaneczky, Anal. Chem. 65:2338-2345 [1993] and Szabo et al. Curr. Opin. Struct. Biol. 5:699-705 [1995]). "Surface plasmon resonance" or "BLA" detects biospecific interactions in real time, without labeling any of the interactants (e.g., BLAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal that can be used as an indication of real-time reactions between biological molecules.

In one embodiment, the target gene product or the test substance is anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid phase can be detected at the end of the reaction. Preferably, the target gene product can be anchored onto a solid surface, and the test compound, (which is not anchored), can be labeled, either directly or indirectly, with detectable labels discussed herein.
It may be desirable to immobilize cancer markers, an anti-cancer marker antibody or its target molecule to facilitate separation of complexes from non-complexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a cancer marker protein, or interaction of a cancer marker protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reagents. Examples of such vessels include microtitrator plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase-cancer marker fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione Sepharose beads (Sigma Chemical, St. Louis, Mo.) or glutathione-de- derivatized microtitrator plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or cancer marker protein, and the mixture incubated under conditions conducive for complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitrator plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above.

Alternatively, the complexes can be dissociated from the matrix, and the level of cancer markers binding or activity determined using standard techniques. Other techniques for immobilizing either cancer marker protein or a target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated cancer marker protein or target molecules can be prepared from biotin-NHS(N-hydroxysuccinimide) using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical).

In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-IgG antibody).

This assay is performed utilizing antibodies reactive with cancer marker protein or target molecules but which do not interfere with binding of the cancer markers protein to its target molecule. Such antibodies can be derivatized to the wells of the plate, and unbound target or cancer markers protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the cancer marker protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the cancer marker protein or target molecule.

Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from unreacted components, by any of a number of standard techniques, including, but not limited to: differential centrifugation (see, for example, Rivas and Minton, Trends Biochem Sci 18:284-7 [1993]); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (see, e.g., Ausubel et al., eds. Current Protocols in Molecular Biology 1999, J. Wiley: New York); and immunoprecipitation (see, for example, Ausubel et al., eds. Current Protocols in Molecular Biology 1999, J. Wiley: New York). Such resins and chromatographic techniques are known to one skilled in the art (See e.g., Heegaard J. Mol. Recognit. 11:141-8 [1998]; Hague and Tweed J. Chromatogr. Biomed. Sci. Appl 699:499-525 [1997]). Further, fluorescence energy transfer may also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

The assay can include contacting the cancer markers protein, mRNA, or biologically active portion thereof with a known compound that binds the cancer marker to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a cancer marker protein or mRNA, wherein determining the ability of the test compound to interact with a cancer marker protein or mRNA includes determining the ability of the test compound to preferentially bind to cancer markers or biologically active portion thereof, or to modulate the activity of a target molecule, as compared to the known compound.

To the extent that cancer markers can, in vivo, interact with one or more cellular or extracellular macromolecules, such as proteins, inhibitors of such an interaction are useful. A homogeneous assay can be used to identify inhibitors.

For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared such that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (see, e.g., U.S. Pat. No. 5,283,317; Zervos et al., Cell 72:223-232 [1993]; Madura et al., J. Biol. Chem. 268:12046-12054 [1993]; Bartel et al., Biotechniques 14:920-924 [1993]; Iwabuchi et al., Oncogene 8:1693-1696 [1993]; and Brent WO 94/10600; each of which is herein incorporated by reference), to identify other proteins, that bind to or interact with cancer markers (“cancer marker-binding proteins” or “cancer marker-bp”) and are involved in cancer marker activity. Such cancer marker-bps can be activators or inhibitors of signals by the cancer marker proteins or targets as, for example, downstream elements of a cancer markers-mediated signaling pathway.

Modulators of cancer markers expression can also be identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of cancer marker mRNA or protein evaluated relative to the level of expression of cancer marker mRNA or protein in the absence of the candidate compound. When expression of cancer marker mRNA or protein is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of cancer marker mRNA or protein expression. Alternatively, when expression of cancer marker mRNA or protein is less (i.e., statistically significantly less) in the presence of the candidate compound than in
its absence, the candidate compound is identified as an inhibitor of cancer marker mRNA or protein expression. The level of cancer markers mRNA or protein expression can be determined by methods described herein for detecting cancer markers mRNA or protein.

A modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a cancer markers protein can be confirmed in vivo, e.g., in an animal such as an animal model for a disease (e.g., an animal with prostate cancer or metastatic prostate cancer; or an animal harboring a xenograft of a prostate cancer from an animal (e.g., human) or cells from a cancer resulting from metastasis of a prostate cancer (e.g., to a lymph node, bone, or liver), or cells from a prostate cancer cell line.

This invention further pertains to novel agents identified by the above-described screening assays (See e.g., below diagnostic-specific therapeutic agents). Accordingly, it is within the scope of this invention to further use an agent identified as described herein (e.g., a cancer marker modulating agent, an antisense cancer marker nucleic acid molecule, a siRNA molecule, a cancer marker specific antibody, or a cancer marker-binding partner) in an appropriate animal model (such as those described herein) to determine the efficacy, toxicity, side effects, or mechanism of action, of treatment with such an agent. Furthermore, novel agents identified by the above-described screening assays can be, e.g., used for treatments as described herein.

VI. Therapeutic Applications

In some embodiments, the present invention provides therapies for cancer (e.g., prostate cancer). In some embodiments, therapies directly or indirectly target cancer markers (e.g., including but not limited to, ERG, ETV1, and ETV4 gene fusions with TMPRSS2).

A. RNA Interference and Antisense Therapies

In some embodiments, the present invention targets the expression of cancer markers. For example, in some embodiments, the present invention employs compositions comprising oligomeric antisense or RNAi compounds, particularly oligonucleotides (e.g., those identified in the drug screening methods described above), for use in modulating the function of nucleic acid molecules encoding cancer markers of the present invention, ultimately modulating the amount of cancer marker expressed.

1. RNA Interference (RNAi)

In some embodiments, RNAi is utilized to inhibit fusion protein function. RNAi represents an evolutionary conserved cellular defense for controlling the expression of foreign genes in most eukaryotes, including humans. RNAi is typically triggered by double-stranded RNA (dsRNA) and causes sequence-specific mRNA degradation of single-stranded target RNAs homologous in response to dsRNA. The mediators of mRNA degradation are small interfering RNA duplexes (siRNAs), which are normally produced from long dsRNA by enzymatic cleavage in the cell. siRNAs are generally approximately twenty-one nucleotides in length (e.g., 21-23 nucleotides in length), and have a base-paired structure characterized by two nucleotide 3′-overhangs. Following the introduction of a small RNA, or RNAi, into the cell, it is believed the sequence is delivered to an enzyme complex called RISC (RNA-induced silencing complex). RISC recognizes the target and cleaves it with an endonuclease. It is noted that if larger RNA sequences are delivered to a cell, RNase III enzyme (Dicer) converts longer dsRNA into 21-23 nt ds siRNA fragments. In some embodiments, RNAi oligonucleotides are designed to target the junction region of fusion proteins.

Chemically synthesized siRNAs have become powerful reagents for genome-wide analysis of mammalian gene function in cultured somatic cells. Beyond their value for validation of gene function, siRNAs also hold great potential as gene-specific therapeutic agents (Tuschl and Borkhardt, Molecular Intervent. 2002; 2(3):158-67, herein incorporated by reference).

The transfection of siRNAs into animal cells results in the potent, long-lasting post-transcriptional silencing of specific genes (Caplen et al, Proc Natl Acad Sci U.S.A. 2001; 98: 9742-7; Elbashir et al., Nature. 2001; 411:494-8; Elbashir et al., Gene Dev. 2001; 15: 188-200; and Elbashir et al., EMBO J. 2001; 20: 6877-88, all of which are herein incorporated by reference). Methods and compositions for performing RNAi with siRNAs are described, for example, in U.S. Pat. No. 6,506,559, herein incorporated by reference.

siRNAs are extraordinarily effective at lowering the amounts of targeted RNA, and by extension proteins, frequently to undetectable levels. The silencing effect can last several months, and is extraordinarily specific, because one nucleotide mismatch between the target RNA and the central region of the siRNA is frequently sufficient to prevent silencing (Brummelkamp et al, Science 2002; 296:550-3; and Helen et al, Nucleic Acids Res. 2002; 30:1757-66, both of which are herein incorporated by reference). An important factor in the design of siRNAs is the presence of accessible sites for siRNA binding. Bahl et al., (J. Biol. Chem., 2003; 278: 15991-15997; herein incorporated by reference) describe the use of a type of DNA array called a scanning array to find accessible sites in mRNAs for designing effective siRNAs. These arrays comprise oligonucleotides ranging in size from monomers to a certain maximum, usually Corners, synthesized using a physical barrier (mask) by stepwise addition of each base in the sequence. Thus the arrays represent a full oligonucleotide complement of a region of the target gene. Hybridization of the target mRNA to these arrays provides an exhaustive accessibility profile of this region of the target mRNA. Such data are useful in the design of antisense oligonucleotides (ranging from 7mers to 25mers), where it is important to achieve a compromise between oligonucleotide length and binding affinity, to retain efficacy and target specificity (Solaim et al, Nucleic Acids Res., 2001; 29(10): 2041-2045). Additional methods and concerns for selecting siRNAs are described for example, in WO 05054270, WO05030854A1, WO05070966A2, J Mol. Biol. 2005 May 13; 348(4):883-93, J Mol. Biol. 2005 May 13; 348(4):871-81, and Nucleic Acids Res. 2003 Aug. 1; 31(15): 4417-24, each of which is herein incorporated by reference in its entirety. In addition, software (e.g., the MWG online siMAX siRNA design tool) is commercially or publicly available for use in the selection of siRNAs.

2. Antisense

In other embodiments, fusion protein expression is modulated using antisense compounds that specifically hybridize with one or more nucleic acids encoding cancer markers of the present invention. The specific hybridization of an oligomeric compound with its target nucleic acid interferes with the normal function of the nucleic acid. This modulation of function of a target nucleic acid by compounds that specifically hybridize to it is generally referred to as “antisense.” The functions of DNA to be interfered with include replication and transcription. The functions of RNA to be interfered with include all vital functions such as, for example, translocation of the RNA to the site of protein translation, translation of protein from the RNA, splicing of the RNA to yield one or more mRNA species, and catalytic activity that may be engaged in or facilitated by the RNA. The overall effect of
such interference with target nucleic acid function is modulation of the expression of cancer markers of the present invention. In the context of the present invention, "modulation" means either an increase (stimulation) or a decrease (inhibition) in the expression of a gene. For example, expression may be inhibited to potentially prevent tumor proliferation.

It is preferred to target specific nucleic acids for antisense.

"Targeting" an antisense compound to a particular nucleic acid, in the context of the present invention, is a multistep process. The process usually begins with the identification of a nucleic acid sequence whose function is to be modulated. This may be, for example, a cellular gene (or mRNA transcribed from the gene) whose expression is associated with a particular disorder or disease state, or a nucleic acid molecule from an infectious agent. In the present invention, the target is a nucleic acid molecule encoding a cancer marker of the present invention. The targeting process also includes determination of a site or sites within this gene for the antisense interaction to occur such that the desired effect, e.g., detection or modulation of expression of the protein, will result. Within the context of the present invention, a preferred intragenic site is the region encompassing the translation initiation or termination codon of the open reading frame (ORF) of the gene. Since the translation initiation codon is typically 5'-AUG (in transcribed mRNA molecules; 5'-ATG in the corresponding DNA molecule), the translation initiation codon is also referred to as the "AUG codon," "the start codon" or the "AUG start codon." A minority of genes have a translation initiation codon having the RNA sequence 5'-GUG, 5'-UUG or 5'-CUU, and 5'-UUA, 5'-ACG and 5'-CUU have been shown to function in vivo. Thus, the terms "translation initiation codon" and "start codon" can encompass many codon sequences, even though the initiator amino acid in each instance is typically methionine (in eukaryotes) or formylmethionine (in prokaryotes). Eukaryotic and prokaryotic genes may have two or more alternative start codons, any one of which may be preferentially utilized for translation initiation in a particular cell type or tissue, or under a particular set of conditions. In the context of the present invention, "start codon" and "translation initiation codon" refer to the codon or codons that are used in vivo to initiate translation of an mRNA molecule transcribed from a gene encoding a tumor antigen of the present invention, regardless of the sequence(s) of such codons.

Translation termination codon (or "stop codon") of a gene may have one of three sequences (i.e., 5'-UAA, 5'-UAG and 5'-UGA; the corresponding DNA sequences are 5'-TAA, 5'-TAG and 5'-TGA, respectively). The terms "start codon region" and "translation initiation codon region" refer to a portion of such an mRNA or gene that encompasses from about 25 to about 50 contiguous nucleotides in either direction (i.e., 5' or 3') from a translation initiation codon. Similarly, the terms "stop codon region" and "translation termination codon region" refer to a portion of such an mRNA or gene that encompasses from about 25 to about 50 contiguous nucleotides in either direction (i.e., 5' or 3') from a translation termination codon.

The open reading frame (ORF) or "coding region," which refers to the region between the translation initiation codon and the translation termination codon, is also a region that may be targeted effectively. Other target regions include the 5' untranslated region (5' UTR), referring to the portion of an mRNA in the 5' direction from the translation initiation codon, and thus including nucleotides between the 5' cap site and the translation initiation codon of an mRNA or corresponding nucleotides on the gene, and the 3' untranslated region (3' UTR), referring to the portion of an mRNA in the 3' direction from the translation termination codon, and thus including nucleotides between the translation termination codon and 3' end of an mRNA or corresponding nucleotides on the gene. The 5' cap of an mRNA comprises an N7-methylated guanosine residue joined to the 5' most residue of the mRNA via a 5'--3' triphosphate linkage. The 5' cap region of an mRNA is considered to include the 5' cap structure itself as well as the first 50 nucleotides adjacent to the cap. The cap region may also be a preferred target region.

Although some eukaryotic mRNA transcripts are directly translated, many contain one or more regions, known as "introns," that are excised from a transcript before it is translated. The remaining (and therefore translated) regions are known as "exons" and are spliced together to form a continuous mRNA sequence. mRNA splice sites (i.e., intron-exon junctions) may also be preferred target regions, and are particularly useful in situations where aberrant splicing is implicated in disease, or where an overproduction of a particular mRNA splice product is implicated in disease. Aberrant fusion junctions due to rearrangements or deletions are also preferred targets. It has also been found that introns can also be effective, and therefore preferred, target regions for antisense compounds targeted, for example, to DNA or pre-mRNA.

In some embodiments, target sites for antisense inhibition are identified using commercially available software programs (e.g., Biognostik, Gottingen, Germany; SysArris Software, Bangalore, India; Antisense Research Group, University of Liverpool, Liverpool, England; GeneTrove, Carlsbad, Calif.). In other embodiments, target sites for antisense inhibition are identified using the accessible site method described in PCT Publ. No. W00198537A2, herein incorporated by reference.

Once one or more target sites have been identified, oligonucleotides are chosen that are sufficiently complementary to the target (i.e., hybridize sufficiently well and with sufficient specificity) to give the desired effect. For example, in preferred embodiments of the present invention, antisense oligonucleotides are targeted to or near the start codon.

In the context of this invention, "hybridization," with respect to antisense compositions and methods, means hydrogen bonding, which may be Watson-Crick, Hoogsteen or reversed Hoogsteen hydrogen bonding, between complementary nucleoside or nucleotide bases. For example, adenine and thymine are complementary nucleobases that pair through the formation of hydrogen bonds. It is understood that the sequence of an antisense compound need not be 100% complementary to that of its target nucleic acid to be specifically hybridizable. An antisense compound is specifically hybridizable when binding of the compound to the target DNA or RNA molecule interferes with the normal function of the target DNA or RNA to cause a loss of utility, and there is a sufficient degree of complementarity to avoid non-specific binding of the antisense compound to non-target sequences under conditions in which specific binding is desired (i.e., under physiological conditions in the case of in vivo assays or therapeutic treatment, and in the case of in vitro assays, under conditions in which the assays are performed).

Antisense compounds are commonly used as research reagents and diagnostics. For example, antisense oligonucleotides, which are able to inhibit gene expression with specificity, can be used to elucidate the function of particular genes. Antisense compounds are also used, for example, to distinguish between functions of various members of a biological pathway.
The specificity and sensitivity of antisense is also applied for therapeutic uses. For example, antisense oligonucleotides have been employed as therapeutic moieties in the treatment of disease states in animals and man. Antisense oligonucleotides have been safely and effectively administered to humans and numerous clinical trials are presently underway. It is thus established that oligonucleotides are useful therapeutic modalities that can be configured to be useful in treatment regimes for treatment of cells, tissues, and animals, especially humans.

While antisense oligonucleotides are a preferred form of antisense compound, the present invention comprehends other oligomeric antisense compounds, including but not limited to oligonucleotide mimetics such as are described below. The antisense compounds in accordance with this invention preferably comprise from about 8 to about 30 nucleobases (i.e., from about 8 to about 30 linked bases), although longer and shorter sequences may find use with the present invention. Preferably particular antisense compounds are antisense oligonucleotides, even more preferably those comprising from about 12 to about 25 nucleobases.

Specific examples of preferred antisense compounds useful with the present invention include oligonucleotides containing modified backbones or non-natural internucleoside linkages. As defined in this specification, oligonucleotides having modified backbones include those that retain a phosphorus atom in the backbone and those that do not have a phosphorus atom in the backbone. For the purposes of this specification, modified oligonucleotides that do not have a phosphorus atom in their internucleoside backbone can also be considered to be oligonucleosides.

Preferred modified oligonucleotide backbones include, for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphorothriesters, aminooxyphosphorothriesters, methyl and other alkyl phosphonates including 3'-alkylphosphonates and chiral phosphonates, phosphonates, phosphoramidates including 3'-amino phosphoramidate and aminooxyphosphorosamidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkoxyphosphorothriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleo-

sides are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Various salts, mixed salts and free acid forms are also included.

Preferred modified oligonucleotide backbones that do not include a phosphorus atom therein have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of nucleoside): siloxane backbones; sulfide, sulfone and sulfone backbones; formacetyl and thiocarbamoyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneiminio and methylenehydradrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage (i.e., the backbone) of the nucleotide units are replaced with novel groups. The base units are maintained for hybridization with an appropriate nucleic acid target compound. One such oligomeric compound, an oligonucleotide mimetic that has been shown to have excellent hybridization properties, is referred to as a peptide nucleic acid (PNA). In PNA compounds, the sugar backbone of an oligonucleotide is replaced with an amide containing backbone, in particular an aminooethylglycine backbone. The nucleobases are retained and are bound directly or indirectly to amine nitrogen atoms of the amide portion of the backbone. Representative United States patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Pat. Nos. 5,539,802; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Further teaching of PNA compounds can be found in Nielsen et al., Science 254:1497 (1991).

Most preferred embodiments of the invention are oligonucleotides with phosphorothioate backbones and oligo-nucleosides with heteroatom backbones, and in particular

\[ \text{CH}_2\text{N(OH)}\text{O} \text{CH}_3 \]

[known as a methylene (methylene) or MM1 backbone],

\[ \text{CH}_2\text{O} \text{N(CH}_3)_2 \text{CH}_2\text{N(OH)}\text{O} \text{CH}_3 \]

[wherein the native phosphodiester backbone is represented as \( \text{O} \rightarrow \text{P} \rightarrow \text{O} \rightarrow \text{CH}_2 \cdot \)]

of the above-referenced U.S. Pat. No. 5,489,677, and the amide backbones of the above-referenced U.S. Pat. No. 5,602,240. Also preferred are oligonucleotides having morpholino and thiono structures of the above-referenced U.S. Pat. No. 5,034,506.

Modified oligonucleotides may also contain one or more substituted sugar moieties. Preferred oligonucleotides comprise one of the following at the 2' position: OH; F; O—S—, or N-alkyl; O—S—, or N-alkenyl; O—S—, or N-alkenyl; or alkyl-O-alkyl, wherein the alkyl, alkenyl and alkylpen may be substituted or unsubstituted C₃ to C₆ alkyl or C₆ to C₁₀ alkyl and alkynyl. Particularly preferred are O(CH₃)₃, O₃CH₃, O(CH₃)₁OH, O(CH₃)₂OH, O(CH₃)₃OH, O(CH₃)₄ONH₂, and O(CH₃)₅ONH₂, where n and m are from 1 to about 10. Other preferred oligonucleotides comprise one of the following at the 2' position: C₆ to C₁₀ lower alkyl, substituted lower alkyl, alkaryl, or alkaryl-alkyl, or alkaryl or O-alkaryl, SO₃CH₃, O₃CH₃, ONH₂, or N₃CH₃, or N₃, N₃H, O₃, heterocycloalkyl, heterocycloalkyl, aminooxyalkylaminol, polycyclolamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmaco-kinetic properties of an oligonucleotide, or a group for improving the pharmacodynamic properties of an oligonucleotide, and other substituents having similar properties. A preferred modified inclusion includes 2'-methoxyethoxy (2'-O—CH₂CH₂OCH₃), also known as 2'-O-(2-methoxyethyl) or 2'-MOE (Martin et al., Helv. Chim. Acta 78:486 [1995]) i.e., an alkoxalkyloxy group. A further preferred modified inclusion includes 2'-dimethylaminooxyethoxy (i.e., a O(CH₃)₂ON (CH₃)₂ group), also known as 2'-DMAOE, and 2'-dimethylaminoethoxy (also known in the art as 2'-O-dimethylaminoethoxy or 2'-DMAOE), i.e., 2'-O—CH₂—N(CH₃)₂.

Other preferred modifications include 2'-methoxy(2'-O—CH₃), 2'-aminoproxy(2'-OCH₂CH₂NH₂) and 2'-fluoro (2'-F). Similar modifications may also be made at other positions on the oligonucleotide, particularly the 3' position of the sugar on the 3' terminal nucleotide or in 2'-5' linked oligonucleotides and the 5' position of 5' terminal nucleotide. Oligonucleotides may also have sugar mimetics such as cyclobutyl moieties in place of the pentofuranosyl sugar.

Oligonucleotides may also include nucleobase (often referred to in the art simply as “base”) modifications or substitutions. As used herein, “unmodified” or “natural” nucleobases include the purine bases adenine (A) and guanine (G), and the pyrimidine bases thymine (T), cytosine (C) and uracil (U). Modified nucleobases include other synthetic and natural nucleobases such as 5-methylcytosine (5-me-C), 5-hy-
droxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiouridine and 2-thiocytosine, 5-halo uracil and cytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amine, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 8-aza guanine and 8-aza adenine, 7-deazaguanine and 7-deaza adenine and 3-deaza guanine and 3-deaza adenine. Further nucleobases include those disclosed in U.S. Pat. No. 3,687,808. Certain of these nucleobases are particularly useful for increasing the binding affinity of the oligomeric compounds of the invention. These include 5-substituted pyrimidines, 6-aza pyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyl adenine, 5-propynyl uracil and 5-propynyl cytosine. 5-methyl cytosine substitutions have been shown to increase nuclear acid duplex stability by 0.6-1.2°C. and are presently preferred base substitutions, even more particularly when combined with 2'-O-methoxymethyl sugar modifications.

Another modification of the oligonucleotides of the present invention involves chemically linking to the oligonucleotide one or more moieties or conjugates that enhance the activity, cellular distribution or cellular uptake of the oligonucleotide. Such moieties include but are not limited to lipid moieties such as a cholesterol moiety, cholic acid, a thioether, (e.g., hexyl-5-tritylthiolo), a thiochelate, an aliphatic chain, (e.g., dodecyl or undecyl residues), a phospholipid, (e.g., di-hexadecyl-rac-glycerol or trithylammonium 1,2-di-O-hexadecyl-rac-glycerol-3-H-phosphonate), a polyanion or a polyethylene glycol chain or adamantane acetic acid, a palmitoyl moiety, or an octadecylamine or hexylamino-carbonyl-oxycarboxy cholesterol moiety.

One skill in the relevant art knows well how to generate oligonucleotides containing the above-described modifications. The present invention is not limited to the antisense oligonucleotides described above. Any suitable modification or substitution may be utilized.

It is not necessary for all positions in a given compound to be uniformly modified, and in fact more than one of the aforementioned modifications may be incorporated in a single compound or even at a single nucleoside within an oligonucleotide. The present invention also includes antisense compounds that are chimeric compounds. “Chimeric” antisense compounds or “chimeras,” in the context of the present invention, are antisense compounds, particularly oligonucleotides, which contain two or more chemically distinct regions, each made up of at least one monomer unit, i.e., a nucleotide in the case of an oligonucleotide compound. These oligonucleotides typically contain at least one region wherein the oligonucleotide is modified so as to confer upon the oligonucleotide increased resistance to nuclease degradation, increased cellular uptake, and/or increased binding affinity for the target nucleic acid. An additional region of the oligonucleotide may serve as a substrate for enzymes capable of cleaving RNA:DNA or RNA:RNA hybrids. By way of example, RNase H is a cellular endonuclease that cleaves the RNA strand of an RNA:DNA duplex. Activation of RNase H, therefore, results in cleavage of the RNA target, thereby greatly enhancing the efficiency of oligonucleotide inhibition of gene expression. Consequently, comparable results can often be obtained with shorter oligonucleotides when chimeric oligonucleotides are used, compared to phosphorothioate deoxyoligonucleotides hybridizing to the same target region. Cleavage of the RNA target can be routinely detected by gel electrophoresis and, if necessary, associated nucleic acid hybridization techniques known in the art.

Chimeric antisense compounds of the present invention may be formed as composite structures of two or more oligonucleotides, modified oligonucleotides, oligonucleosides and/or oligonucleotide mimetics as described above.

The present invention also includes pharmaceutical compositions and formulations that include the antisense components of the present invention as described below.

B. Gene Therapy

The present invention contemplates the use of any genetic manipulation for use in modulating the expression of cancer markers of the present invention. Examples of genetic manipulation include, but are not limited to, gene knockout (e.g., removing the fusion gene from the chromosome using, for example, recombinase), expression of antisense constructs with or without inducible promoters, and the like. Delivery of nucleic acid constructs to cells in vitro or in vivo may be conducted using any suitable method. A suitable method is one that introduces the nucleic acid construct into the cell such that the desired event occurs (e.g., expression of an antisense construct). Genetic therapy may also be used to deliver siRNA or other interfering molecules that are expressed in vivo (e.g., upon stimulation by an inducible promoter (e.g., an androgen-responsive promoter)).

Introduction of molecules carrying genetic information into cells is achieved by any of various methods including, but not limited to, directed injection of naked DNA constructs, bombardment with gold particles loaded with said constructs, and macromolecule mediated gene transfer using, for example, liposomes, biopolymers, and the like. Preferred methods use gene delivery vehicles derived from viruses, including, but not limited to, adenoviruses, retroviruses, vaccinia viruses, and adeo-associated viruses. Because of the higher efficiency as compared to retroviruses, vectors derived from adenoviruses are the preferred gene delivery vehicles for transferring nucleic acid molecules into host cells in vivo. Adenoviral vectors have been shown to provide very efficient in vivo gene transfer into a variety of solid tumors in animal models and into human solid tumor xenografts in immune-deficient mice. Examples of adenoviral vectors and methods for gene transfer are described in PCT publications WO 00/17238 and WO 00/09675 and U.S. Pat. Nos. 6,033,908, 6,019,978, 6,001,557, 5,994,132, 5,994,128, 5,994,106, 5,981,225, 5,885,808, 5,872,154, 5,830,730, and 5,824,544, each of which is herein incorporated by reference in its entirety.

Vectors may be administered to subject in a variety of ways. For example, in some embodiments of the present invention, vectors are administered into tumors or tissue associated with tumors using direct injection. In other embodiments, administration is via the blood or lymphatic circulatory system (See e.g., PCT publication 99/02,685 herein incorporated by reference in its entirety). Exemplary dose levels of adenoviral vector are preferably 10⁹ to 10¹¹ vector particles added to the perfusate.

C. Antibody Therapy

In some embodiments, the present invention provides antibodies that target prostate tumors that express a cancer marker of the present invention, e.g., ERG, ETV1, or ETV4 fusions with TMPRSS2. Any suitable antibody (e.g., monoclonal, polyclonal, or synthetic) may be utilized in the therapeutic methods disclosed herein. In preferred embodiments, the antibodies used for cancer therapy are humanized antibodies. Methods for humanizing antibodies are well known in
the art (See e.g., U.S. Pat. Nos. 6,180,370, 5,585,089, 6,054, 297, and 5,565,332; each of which is herein incorporated by reference).

In some embodiments, the therapeutic antibodies comprise an antibody generated against a cancer marker of the present invention (e.g., ERG, ETV1, or ETV4 fusions with TMRPSS2), wherein the antibody is conjugated to a cytotoxic agent. In such embodiments, a tumor specific therapeutic agent is generated that does not target normal cells, thus reducing many of the detrimental side effects of traditional chemotherapy. For certain applications, it is envisioned that the therapeutic agents will be pharmacological agents that will serve as useful agents for attachment to antibodies, particularly cytotoxic or otherwise antineutal agents having the ability to kill or suppress the growth of cell division of endo
telial cells. The present invention contemplates the use of any pharmacological agent that can be conjugated to an antibody, and delivered selectively to the tumor. Examples of antineural agents
include chemotherapeutic agents, radioisotopes, and cytotoxins. The therapeutic antibodies of the present invention may include a variety of cytotoxic moieties, including but not limited to, radioactive isotopes (e.g., iodine-131, iodine-123, technetium-99m, indium-111, rhenium-188, rhenium-186, gallium-67, copper-67, yttrium-90, iodine-125 or astutine-211), hormones such as a steroid, antineumabibolites such as cytokines (e.g., arabinoside, fluorouracil, methotrexate or aminopterin; an anthracycline; mitomycinc C), vinca alkaloids (e.g., demecolcine; etoposide; mithramycin), and antineuralkalyting agent such as chlorambucil or melphalan. Other embodiments may include agents such as a coagulant, a cytokine, growth factor, bacterial endotoxin or the lipid A moiety of bacterial endotoxins. For example, in some embodiments, mutagens may include plant-, fungi- or bacteria-derived toxin, such as an A chain toxins, a ribosome inactivating protein, ct-sarcin, aspergillarin, restrictocin, a ribo
nuclelease, diphtheria toxin or pseudomonas exotoxin, to mention just a few examples. In some preferred embodiments, deglycosylated ricin A chain is utilized.

In any event, it is proposed that agents such as these may, if desired, be successfully conjugated to an antibody, in a manner that will allow their targeting, internalization, release or presentation to blood components at the site of the targeted tumor cells as required using known conjugation technology (See, e.g., Ghose et al., Methods Enzymol., 93:280 [1983]).

For example, in some embodiments the present invention provides immunoconjugates targeted a cancer marker of the present invention (e.g., ERG or ETV1 fusions). Immunotoxins are conjugates of a specific targeting agent typically a tumor-directed antibody or fragment, with a cytotoxic agent, such as a toxin moiety. The targeting agent directs the toxin to, and thereby selectively kills, cells expressing the targeted antigen. In some embodiments, therapeutic antibodies employ crosslinkers that provide high in vivo stability (Thorpe et al., Cancer Res., 48:6396 [1988]).

In other embodiments, particularly those involving treatment of solid tumors, antibodies are designed to have a cytotoxic or otherwise antineural effect against the tumor vasculature, by suppressing the growth of cell division of the vascular endothelial cells. This attack is intended to lead to a tumor-localized vascular collapse, depriving the tumor cells, particularly those tumor cells distal of the vasculature, of oxygen and nutrients, ultimately leading to cell death and tumor necrosis.

In preferred embodiments, antibody based therapies are formulated as pharmaceutical compositions as described below. In preferred embodiments, administration of an anti
body composition of the present invention results in a measurable decrease in cancer (e.g., decrease or elimination of tumor).

D. Pharmaceutical Compositions

The present invention further provides pharmaceutical compositions (e.g., comprising pharmaceutical agents that modulate the expression or activity of gene fusions of the present invention). The pharmaceutical compositions of the present invention may be administered in a number of ways depending upon whether local or systemic treatment is desired and upon the area to be treated. Administration may be topical (including ophthalmic and to mucous membranes including vaginal and rectal delivery), orally or parenteral. Parenteral administration includes intravenous, intraarterial, intracutaneous, intraperitoneal or intramuscular injection or infusion; or intracranial, e.g., intrathecal or intraventricular, administration.

Pharmaceutical compositions and formulations for topical administration may include transdermal patches, ointments, lotions, creams, gels, drops, suppositories, sprays, liquids and powders. Conventional pharmaceutical carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable.

Compositions and formulations for oral administration include powders or granules, suspensions or solutions in water or non-aqueous media, capsules, sachets or tablets. Thickeners, flavoring agents, diluents, emulsifiers, dispersing aids or binders may be desirable.

Compositions and formulations for parenteral, intrathecal or intraventricular administration may include sterile aque
ous solutions that may also contain buffers, diluents and other suitable additives such as, but not limited to, penetration enhancers, carrier compounds and other pharmaceutically acceptable carriers or excipients.

Pharmaceutical compositions of the present invention include, but are not limited to, solutions, emulsions, and liposome-containing formulations. These compositions may be prepared from a variety of components that include, but are not limited to, preformed liquids, self-emulsifying solids and self-emulsifying semisolids.

The pharmaceutical compositions of the present invention, which may conveniently be presented in unit dosage form, may be prepared according to conventional techniques well known in the pharmaceutical industry. Such techniques include the step of bringing into association the active ingredients with the pharmaceutical carrier(s) or excipient(s). In general the formulations are prepared by uniformly and inti
mately bringing into association the active ingredients with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product.

The compositions of the present invention may be formulated into any of many possible dosage forms such as, but not limited to, tablets, capsules, liquid syrups, soft gels, supposito
tories, and enemas. The compositions of the present invention may also be formulated as suspensions in aqueous, non-
aqueous or mixed media. Aqueous suspensions may contain substances that increase the viscosity of the suspen
sion including, for example, sodium carboxymethylcellulose, sorbitol and/or dextran. The suspension may also contain stabilizers.

In one embodiment of the present invention the pharmaceutical compositions may be formulated as solutions. Pharmaceutical compositions may include formulations such as, but not limited to, emulsions, microemulsions, creams, gels, and
luposomes. While basically similar in nature these formulations vary in the components and the consistency of the final product.

Agents that enhance uptake of oligonucleotides at the cellular level may also be added to the pharmaceutical and other compositions of the present invention. For example, cationic lipids, such as lipofectin (U.S. Pat. No. 5,705,188), cationic glycerol derivatives, and polycationic molecules, such as polylysine (WO 97/30731), also enhance the cellular uptake of oligonucleotides.

The compositions of the present invention may additionally contain other adjunct components conventionally found in pharmaceutical compositions. Thus, for example, the compositions may contain additional, compatible, pharmaceutically-active materials such as, for example, antipruritics, astringents, local anesthetics or anti-inflammatory agents, or may contain additional materials useful in physically formulating various dosage forms of the compositions of the present invention, such as dyes, flavoring agents, preservatives, antioxidants, opacifiers, thickening agents and stabilizers. However, such materials, when added, should not unduly interfere with the biological activities of the components of the compositions of the present invention. The formulations can be sterilized and, if desired, mixed with auxiliary agents, e.g., lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic pressure, buffers, colorings, flavorings and/or aromatic substances and the like which do not deleteriously interact with the nucleic acid(s) of the formulation.

Certain embodiments of the invention provide pharmaceutical compositions containing (a) one or more antisense compounds and (b) one or more other chemotherapeutic agents that function by a non-antisense mechanism. Examples of such chemotherapeutic agents include, but are not limited to, antibiotics such as daunorubicin, actinomycin, doxorubicin, bleomycin, mitomycin, nitrogen mustard, chlorambucil, melphalan, cyclophosphamide, 6-mercaptopurine, 6-thioguanine, cytarabine (CA), 5-fluouracil (FU), fluorouridine (5-FUdR), methotrexate (MTX), colchicine, vincristine, vinblastine, etoposide, teniposide, cisplatin and diethylstilbestrol (DES). Anti-inflammatory drugs, including but not limited to nonsteroidal anti-inflammatory drugs and corticosteroids, and antiviral drugs, including but not limited to ribavirin, vidarabine, acyclovir and ganciclovir, may also be combined in compositions of the invention. Other non-antisense chemotherapeutic agents are also within the scope of this invention. Two or more combined compounds may be used together or sequentially.

Dosing is dependent on severity and responsiveness of the disease state to be treated, with the course of treatment lasting from several days to several months, or until a cure is effected or a diminution of the disease state is achieved. Optimal dosing schedules can be calculated from measurements of drug accumulation in the body of the patient. The administering physician can easily determine optimum dosages, dosing methodologies and repetition rates. Optimum dosages may vary depending on the relative potency of individual oligonucleotides, and can generally be estimated based on EC50's found to be effective in vitro and in vivo animal models or based on the examples described herein. In general, dosage is from 0.01 μg to 100 g per kg of body weight, and may be given once or more daily, weekly, monthly or yearly. The treating physician can estimate repetition rates for dosing based on measured residence times and concentrations of the drug in bodily fluids or tissues. Following successful treatment, it may be desirable to have the subject undergo maintenance therapy to prevent the recurrence of the disease state, wherein the oligonucleotide is administered in maintenance doses, ranging from 0.01 μg to 100 g per kg of body weight, once or more daily, to once every 20 years.

VII. Transgenic Animals

The present invention contemplates the generation of transgenic animals comprising an exogenous cancer marker gene (e.g., gene fusion) of the present invention or variants thereof (e.g., truncations or single nucleotide polymorphisms). In preferred embodiments, the transgenic animal displays an altered phenotype (e.g., increased or decreased presence of markers) as compared to wild-type animals. Methods for analyzing the presence or absence of such phenotypes include but are not limited to, those disclosed herein. In some preferred embodiments, the transgenic animals further display an increased or decreased growth of tumors or evidence of cancer.

The transgenic animals of the present invention find use in drug (e.g., cancer therapy) screens. In some embodiments, test compounds (e.g., a drug that is suspected of being useful to treat cancer) and control compounds (e.g., a placebo) are administered to the transgenic animals and the control animals and the effects evaluated.

The transgenic animals can be generated via a variety of methods. In some embodiments, embryonic cells at various developmental stages are used to introduce transgenes for the production of transgenic animals. Different methods are used depending on the stage of development of the embryonal cell. The zygote is the best target for micro-injection. In the mouse, the male pronucleus reaches the size of approximately 20 micrometers in diameter that allows reproducible injection of 1-2 picoliters (pl) of DNA solution. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host genome before the first cleavage (Brinster et al., Proc. Natl. Acad. Sci. USA 82:4438-4442 [1985]). As a consequence, all cells of the transgenic non-human animal will carry the incorporated transgene. This will in general also be reflected in the sufficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene.

U.S. Pat. No. 4,873,191 describes a method for the micro-injection of zygotes; the disclosure of this patent is incorporated herein in its entirety.

In other embodiments, retroviral infection is used to introduce transgenes into a non-human animal. In some embodiments, the retroviral vector is utilized to transfect oocytes by injecting the retroviral vector into the perivitelline space of the oocyte (U.S. Pat. No. 6,080,912, incorporated herein by reference). In other embodiments, the developing non-human embryo can be cultured in vitro to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection (Junenich, Proc. Natl. Acad. Sci. USA 73:1260 [1976]). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (Hogan et al., in Manipulating the Mouse Embryo, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. [1986]). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahn et al., Proc. Natl. Acad. Sci. USA 82:6927 [1985]). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Stewart et al., EMBO J., 6:383 [1987]). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (Jahn et al., Nature 298:623 [1982]). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of cells that form the transgenic animal. Further, the founder may contain various retroviral insertions of the transgene at dif-
different positions in the genome that generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germline, albeit with low efficiency, by intrauterine retroviral infection of the midgestation embryo (Jahner et al., supra [1982]). Additional means of using retroviruses or retroviral vectors to create transgenic animals known to the art involve the micro-injection of retroviral particles or mitomycin C-treated cells producing retrovirus into the perivitelline space of fertilized eggs or early embryos (PCT International Application WO 94/08832 [1990], and Haskell and Bowen, Mol. Reprod. Dev., 40:386 [1995]).

In other embodiments, the transgene is introduced into embryonic stem cells and the transfected stem cells are utilized to form an embryo. ES cells are obtained by culturing pre-implantation embryos in vitro under appropriate conditions (Evans et al., Nature 292:154 [1981]; Bradley et al., Nature 309:255 [1984]; Gossler et al., Proc. Acad. Sci. USA 83:9065 [1986]; and Robertson et al., Nature 322:445 [1986]). Transgenes can be efficiently introduced into the ES cells by DNA transfection by a variety of methods known to the art including calcium phosphate co-precipitation, protoplast or spheroplast fusion, lipofection and DEAE-dextran-mediated transfection. Transgenes may also be introduced into ES cells by retrovirus-mediated transduction or by micro-injection. Such transfected ES cells can thereafter colonize an embryo following their introduction into the blastocoeil of a blastocyst-stage embryo and contribute to the germ line of the resulting chimeric animal (for review, see, Jaenisch, Science 240:1468 [1988]). Prior to the introduction of transfected ES cells into the blastocoeil, the transfected ES cells may be subjected to various selection protocols to enrich for ES cells which have integrated the transgene assuming that the transgene provides a means for such selection. Alternatively, the polymerase chain reaction may be used to screen for ES cells that have integrated the transgene. This technique obviates the need for growth of the transfected ES cells under appropriate selective conditions prior to transfer into the blastocoeil.

In still other embodiments, homologous recombination is utilized to knock-out gene function or create deletion mutants (e.g., truncation mutants). Methods for homologous recombination are described in U.S. Pat. No. 5,614,396, incorporated herein by reference.

EXPERIMENTAL

The following examples are provided in order to demonstrate and further illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

EXAMPLE 1

ERG and ETV1 Gene Fusions

A. Materials and Methods

Cancer Outlier Profile Analysis (COPA)

COPA analysis was performed on 132 gene expression data sets in Oncomine 3.0 comprising 10,486 microarray experiments. In addition, data from 99 amplified laser-capture microdissected prostate tissue samples were included in the COPA analysis. COPA has three steps. First, gene expression values are median centered, setting each gene's median expression value to zero. Second, the median absolute deviation (MAD) is calculated and scaled to 1 by dividing each gene expression value by its MAD. Median and MAD were used for transformation as opposed to mean and standard deviation so that outlier expression values do not unduly influence the distribution estimates, and are thus preserved post-normalization. Third, the 75th, 90th, and 95th percentiles of the transformed expression values are tabulated for each gene and then genes are rank-ordered by their percentile scores, providing a prioritized list of outlier profiles.

Samples

Tissues utilized were from the radical prostatectomy series at the University of Michigan and from the Rapid Autopsy Program (Shah et al., Cancer Res 64, 9209 (Dec. 15, 2004)), which are both part of University of Michigan Prostate Cancer Specialized Program of Research Excellence (S.P.O.R.E.) Tissue Core.

Tissues were also obtained from a radical prostatectomy series at the University Hospital Ulm (Ulm, Germany). All samples were collected from consented patients with prior institutional review board approval at each respective institution. Total RNA from all samples was isolated with Trizol (Invitrogen) according to the manufacturer's instructions. Total RNA was also isolated from RWPE, PC3, PC3-AR (Dai et al., Steroids 61, 531 (1996)), LNCaP, VCaP and DuCaP cell lines. RNA integrity was verified by denaturing formaldehyde gel electrophoresis or the Agilent Bioanalyzer 2100. A commercially available pool of benign prostate tissue total RNA (CPP, Clontech) was also used.

Quantitative PCR (QPCR)

Quantitative PCR (QPCR) was performed using SYBR Green dye on an Applied Biosystems 7300 Real Time PCR system essentially as described (Chinnaiyan et al., Cancer Res 65, 3328 (2005); Rubin et al., Cancer Res 64, 3814 (2004)). Briefly, 1-5 ng of total RNA was reverse transcribed into cDNA using SuperScript III (Invitrogen) in the presence of random primers or random primers and oligo dT primers. All reactions were performed with SYBR Green Master Mix (Applied Biosystems) and 25 ng of both the forward and reverse primer using the manufacturer's recommended thermocycling conditions. All reactions were subjected to melt curve analysis and products from selected experiments were resolved by electrophoreses on 1.5% agarose gels. For each experiment, threshold levels were set during the exponential phase of the QPCR reaction using Sequence Detection Software version 1.2.2 (Applied Biosystems). The amount of each target gene relative to the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH) for each sample was determined using the comparative threshold cycle (Ct) method (Applied Biosystems User Bulletin #2), with the cDNA sample serving as the calibrator for each experiment described in the figure legend. All oligonucleotide primers were synthesized by Integrated DNA Technologies. GAPDH primers were as described (Vandesompele et al., Genome Biol 3, RESEARCH00534 (2002)) and all other primers are listed (Table 4). Approximately equal efficiencies of the primers were confirmed using serial dilutions of prostate cancer cDNA or plasmid templates in order to use the comparative Ct method.

RNA Ligase Mediated Rapid Amplification of cDNA Ends (RLM-RACE)

RNA ligase mediated rapid amplification of cDNA ends was performed using the GeneRacer RLM-RACE kit (Invitrogen), according to the manufacturer's instructions. Initially, samples were selected based on expression of ERG or ETV1 by QPCR. Five micrograms of total RNA was treated with calf intestinal phosphatase to remove 5' phosphates from truncated mRNA and non-mRNA and decapped with tobacco acid phosphatase. The GeneRacer RNA Oligo was ligated to full length transcripts and reverse transcribed using SuperScript III. To obtain 5' ends, first-strand cDNA was
amplified with Platinum Taq High Fidelity (Invitrogen) using the GeneRacer 5' Primer and ETV1 exon 4-5_r for ETV1 or the GeneRacer 5' Primer and ERG exon 4a_r or ERG exon 4b_r for ERG. Primer sequences are given (Table S2). Products were resolved by electrophoresis on 1.5% agarose gels and bands were excised, purified and TOPO TA cloned into pCR 4-TOPO. Purified plasmid DNA from at least 4 colonies was sequenced bi-directionally using M13 Reverse and M13 Forward (–20) primers or T3 and 17 primers on an ABI Model 3730 automated sequencer by the University of Michigan DNA Sequencing Core. RLM-RACEd cDNA was not used for the other assays.

Reverse-transcription PCR for TMPRSS2:ERG Fusion

After identifying TMPRSS2:ERG positive cases using QPCR as described above, the same cDNA samples were PCR amplified with Platinum Taq High Fidelity and TMPRSS2:ERG primers. Products were resolved by electrophoresis, cloned into pCR 4-TOPO and sequenced as described above.

In Vitro Androgen Responsiveness

RWPE, LNCaP, VCaP DuCaP PC3 and PC3 cells stably transfected with the human androgen receptor (PC3+AR) (3) were treated for 24 h with 1% ethanol control or 1 nM of the synthetic androgen R1881. Total RNA was isolated and subjected to reverse transcription and QPCR as described above with ERG exon 5-6_f and 5-6_r primers. The relative amount of ERG/GAPDH for each sample was calibrated to the RWPE control sample.

Fluorescence In Situ Hybridization (FISH)

Formalin-fixed paraffin-embedded (FFPE) tissue sections from normal peripheral lymphocytes and the metastatic prostate cancer samples MET-26 and MET-28 were used for interphase fluorescence in situ hybridization (FISH) analysis. In addition, interphase FISH was performed on a tissue microarray containing cores from FFPE sections of 13 clinically localized prostate cancer and 16 metastatic prostate cancer samples. A two-color, two-signal approach was employed to evaluate the fusion of TMPRSS2 and ETV1, with probes spanning most of the respective gene loci. The biotin-14-dCTP labeled BAC clone RP11-1241L22 was used for the ETV1 locus and the digoxin-lUTP labeled BAC clone RP11-35CD21 was used for the TMPRSS2 locus. For analyzing gene rearrangements involving ERG, a split-signal probe strategy was used, with two probes spanning the ERG locus (biotin-14-dCTP labeled BAC clone RP11-476D17 and digoxin-lUTP labeled BAC clone RP11-95121). All BAC clones were obtained from the Children’s Hospital of Oakland Research Institute (CHORI). Prior to tissue analysis, the integrity and purity of all probes were verified by hybridization to metaphase spreads of normal peripheral lymphocytes. Tissue hybridization, washing and color detection were performed as described (Rubin et al., Cancer Res 64, 3814 (2004); Garraway et al., Nature 436, 117 (2005)).

B. Results

Cancer Outlier Profile Analysis

In recent years, gene expression profiling with DNA microarrays has become a common method to study the cancer transcriptome. Microarray studies have provided great insight into the molecular heterogeneity of cancer, often identifying novel molecular subtypes of disease that correspond to tumor histology, patient outcome, and treatment response (Valk et al., N Engl J Med 350, 1617 (2004)). However, in general, transcriptome analysis has not led to the discovery of novel causal cancer genes. It was hypothesized that rearrangements and highlevel copy number changes that result in marked over-expression of an oncogene should be evident in transcriptome data, but not necessarily by traditional analytical approaches.

In the majority of cancer types, heterogeneous patterns of oncogene activation have been observed, thus traditional analytical methods that search for common activation of genes across a class of cancer samples (e.g., t-test or signal-to-noise ratio) will fail to find such oncogene expression profiles. Instead, a method that searches for marked over-expression in a subset of cases is needed. Experiments conducted during the course of development of the present invention resulted in the development of Cancer Outlier Profile Analysis (COPA). COPA seeks to accentuate and identify outlier profiles by applying a simple numerical transformation based on the median and median absolute deviation of a gene expression profile (Ross et al., Blood 102, 2951 (2003)). This approach is illustrated in FIG. 5A. COPA was applied to the Oncomine database (Bittner et al., Nature 406, 536 (2000)), which comprised a compendium of 132 gene expression datasets representing 10,486 microarray experiments. COPA correctly identified several outlier profiles for genes in specific cancer types in which a recurrent rearrangement or high-level amplification is known to occur. The analysis was focused on outlier profiles of known causal cancer genes, as defined by the Cancer Gene Census (Vasselli et al., Proc Natl Acad Sci USA 100, 6958 (2003)), that ranked in the top 10 outlier profiles in an Oncomine dataset (Table 1 and Table 3). For example, in the Vakil et al. acute myeloid leukemia (AML) dataset, RUNX1T1 (ETO) had the strongest outlier profile at the 95th percentile, consistent with this gene’s known translocation and oncogenic activity in a subset of AML (Davis et al., Proc Natl Acad Sci USA 100, 6051 (2003)) (Table 1). The outlier profile precisely associated with cases that had a documented (8;21) translocation which fuses RUNX1 (AML1) and RUNX1T1 (ETO) (FIG. 5B). Similarly, in the Ross et al. acute lymphoblastic leukemia (ALL) dataset, PAX5 showed the strongest outlier profile at the 95th percentile, consistent with the E2A-PBX1 translocation known to occur in a subset of ALL (Segal et al., J Clin Oncol 21, 1775 (2003)) (Table 1).

Again, the outlier expression profile perfectly correlated with the characterized t(1;19) E2A-PBX1 translocation in this panel of ALLs (FIG. SIC).

Identification of Outlier Profiles for ETS Family Members ERG and ETV1 in Prostate Cancer

Novel COPA predictions were next examined. In several independent datasets, COPA identified strong outlier profiles in prostate cancer for ERG and ETV1, two ETS family transcription factors that are known to be involved in oncogenic translocations in Ewing’s sarcoma and myeloid leukemias (Lapointe et al., Proc Natl Acad Sci USA 101, 811 (2004); Tian et al., N Engl J Med 349, 2483 (2005)). In the Dharsekar et al. (Keats et al., Blood 105, 4060 (2005)), Welsh et al. (Dharsekar et al., Faseb J 19, 243 (2005)) and Lapoint et al. (Wang et al., Lancet 365, 671 (2005)) prostate cancer gene expression datasets, ERG had the highest scoring outlier profile at the 75th percentile (Table 1), while in the Lapoint et al. and Tomlins et al. (Welsh et al., Cancer Res 61, 5974 (2001)) datasets, ETV1 had the highest scoring outlier profile at the 90th percentile (Table 1). In total, COPA ranked ERG or ETV1 within the top ten outlier genes nine times in seven independent prostate cancer profiling studies. Both ERG and ETV1 are involved in oncogenic translocations in Ewing’s sarcoma. Fusion of the 5’ activation domain of the EWS gene to the highly conserved 3’ DNA binding domain of an ETS family member, such as ERG (t(21;22)(q22;q12) or ETV1 (t(7;22)(p21;q12)), is characteristic of Ewing’s sarcoma (Lapoint et al., supra; Zhan et al., Blood 99, 1745
(2002); Fonseca et al., Cancer Res 64, 1546 (2004)). Because translocations involving ETS family members are functionally redundant in oncogenic transformation, only one type of translocation is typically observed in each case of Ewing’s sarcoma.

It was contemplated that if ERG and ETV1 are similarly involved in the development of prostate cancer, their outlier profiles should be mutually exclusive, that is, each case should over-express only one of the two genes. Mutations in functionally redundant genes, or genes in the same oncogenic pathway, are unlikely to be co-selected for in neoplastic progression. The joint expression profiles of ERG and ETV1 were examined across several prostate cancer datasets and it was found that they showed mutually exclusive outlier profiles. ERG and ETV1 expression profiles from two large-scale transcriptome studies (Wang et al., supra; Cheok et al., Nat Genet. 34, 85 (2003)), which profiled grossly dissected prostate tissues using different microarray platforms, were identified (Fig. 1A, left and middle panels). The study by Lapointe et al. profiled benign prostate tissue, clinically localized prostate cancer, and metastatic prostate cancer, with ERG and ETV1 outlier expression restricted to prostate cancer and metastatic prostate cancer, while the study by Glinsky et al. profiled clinically localized prostate cancers only. In both studies, prostate cancers exclusively expressed ERG or ETV1 (Fig. 1A, right panel). Similar results were found in a profiling study of 99 prostate tissue samples obtained by laser capture microdissection (LCM) (Welsh et al., supra). In addition to exclusive outlier expression of either ERG or ETV1 (Fig. 1B, right panel), results from the LCM study demonstrated that ETV1 and ERG are only over-expressed in epithelial cells from prostate cancer or metastatic prostate cancer, but not in the putative precursor lesion prostatic intraductal neoplasia (PIN) or adjacent benign epithelium. To directly determine whether the observed exclusive outlier pattern is consistent with other translocations where an activating gene can fuse with multiple partners, the Zhan et al. multiple myeloma dataset (Dhanasekaran et al., Nature 412, 822 (2001)) was examined. Recurrent fusions of the immunoglobulin heavy chain promoter to CCND1 or FGFR3, (t11,14) or (4;14) respectively, characterize specific subsets of multiple myeloma (Wigle et al., Cancer Res 62, 3005 (2002)). These translocations were reflected in the outlier profile analysis (Fig. 1C), as CCND1 was the highest scoring outlier at the 75th percentile and FGFR3 was the third highest scoring outlier at the 95th percentile (Table I). Except for two cases, myeloma samples showed exclusive over-expression of CCND1 or FGFR3 (Fig. 1C, right panel). Together, the outlier profiles of ERG and ETV1 across multiple prostate cancer data sets are consistent with other causal mutations in various human malignancies. The exclusive over-expression of ERG or ETV1 in individual prostate cancer samples is consistent with other neoplasms in which an activating gene can fuse with biologically redundant partner genes, such as in multiple myeloma.

Discovery of a Recurrent Gene Fusion of TMPRSS2 to ERG or ETV1 in Prostate Cancer.

The mechanism of ERG and ETV1 over-expression in individual prostate cancer samples was next determined. Prostate cancer cell lines and clinical specimens that over-expressed ERG or ETV1 were identified by performing quantitative PCR (QPCR) (Fig. 2A). The LNCaP prostate cancer cell line and two specimens obtained from a patient who died of hormone refractory metastatic prostate cancer (MET26R), residual primary carcinoma in the prostate and MET26LN, a lymph node metastasis) markedly over-expressed ETV1 by QPCR (Fig. 2A). Five independent metastatic foci from different anatomical locations as well as the residual carcinoma in the prostate from this patient also over-expressed ETV1 by DNA microarray analysis (Welsh et al., supra), suggesting that ETV1 activation occurred in the primary tumor before widespread metastasis. A lymph node metastasis was also identified from a second patient who died of hormone refractory metastatic prostate cancer (MET28LN) and two prostate cancer cell lines, VCaP and DU145, that over-expressed ERG (Fig. 2A). These cell lines were independently isolated from a vertebral metastasis (VCaP) and a dorsal metastasis (DU145) from a third patient with hormone-refractory prostate cancer (Goldberg et al., Science 286, 531 (1999); Rosenwald et al., Cancer Cell 3, 185 (2003)). The common over-expression of ERG in these two cell lines again suggests that ERG activation occurred before widespread metastasis. Taken together, these results suggest that specific genetic events may activate ERG or ETV1 in individual samples during prostate tumorigenesis.

In an effort to characterize these genetic events, samples with high ERG or ETV1 expression were tested for chromosomal amplifications at their respective loci (7p21.2 and 21q22.3). By QPCR on genomic DNA, amplification of ERG or ETV1 in samples with respective transcript over-expression (Soriniou et al., Proc Natl Acad Sci USA 100, 10393 (2003)) was not found. Next, the occurrence of DNA rearrangements was assayed. Because the primers used for the QPCR described above were located 5’ to the known breakpoints for ERG and ETV1 in Ewing’s sarcoma, it was unlikely that the same translocations occur in prostate cancer. Accordingly, the expression level of ETV1 exons was measured by exon-walking QPCR in the samples identified above that displayed ETV1 over-expression. Five primer pairs spanning ETV1 exons 2 through 7 were used and LNCaP cells showed essentially uniform over-expression of all measured ETV1 exons, and both MET26 specimens showed >90% reduction in the expression of ETV1 exons 2 and 3 compared to exons 4-7 (Fig. 2B). Potential explanations for this result include alternative splicing, a novel cancer-specific isoform or an unreported rearrangement.

In order to characterize the full length ETV1 transcript, 5’ RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE) was performed on LNCaP cells and MET26-LN. In addition, RLM-RACE was performed to obtain the full length transcript of ERG in MET28-LN. For PCR amplification of ETV1 from the RLM-RACE cDNA, a forward primer complementary to the RNA-oligonucleotide ligated to the 5’ end of complete transcripts and a reverse primer in exon 4, the 5’-most exon that was over-expressed in both LNCaP cells and MET26-LN was used. Utilizing a similar strategy as described above, it was determined that exon 4 of ERG was over-expressed in MET28-LN. A reverse primer in this exon was utilized for PCR amplification of RLM-RACE cDNA. Sequencing of the cloned products revealed fusions of the prostate specific gene TMPRSS2 (28) (21q22.2) with ETV1 in MET26-LN and with ERG in MET28-LN (Fig. 2C). In MET26-LN, two RLM-RACE PCR products were identified. The first product, TMPRSS2:ETV1a, resulted in a fusion of the complete exon 1 of TMPRSS2 with the beginning of exon 4 of ETV1 (Fig. 2C). The second product, TMPRSS2: ETV1b, resulted in a fusion of exons 1 and 2 of TMPRSS2 with the beginning of exon 4 of ETV1 (Fig. 6). Both products are consistent with the exon-walking QPCR described above, where MET26-LN showed loss of over-expression in exons 2 and 3. In MET28-LN, a single RLM-RACE PCR product was identified and sequencing revealed a fusion of the complete exon 1 of TMPRSS2 with the beginning of exon 4 of ERG (TMPRSS2:ERGα) (Fig. 2C).
Validation of TMPRSS2:ERG and TMPRSS2:ETV1 Gene Fusions in Prostate Cancer

Based on these results, QPCR primer pairs were designed with forward primers in TMPRSS2 and reverse primers in exon 4 of ERG and ETV1. SYBR Green QPCR was performed using both primer pairs across a panel of samples from 42 cases of clinically localized prostate cancer and metastatic prostate cancer, with representative results depicted (FIGS. 2, D and E). These results demonstrate that only samples with high levels of ETV1 or ERG express the respective fusion product with TMPRSS2. Although QPCR resulted in measurable product after 35 cycles in some negative samples, melt curve analysis revealed distinct products in positive and negative samples, and gel electrophoresis of products after the 40 cycle QPCR analysis revealed only primer dimmers in negative fusion samples (FIGS. 2, D and E). The evaluation of primer dimers may in part be explained by the difficulty in designing primers entirely in exon 1 of TMPRSS2 due to the high GC content (80.3%). However, the specific expression of TMPRSS2:ERGα, TMPRSS2:ETV1a and TMPRSS2:ETV1b fusions was confirmed using Taqman QPCR, with the forward primer spanning the respective fusion, and in each case, products were only detected in the same cases as the SYBR Green QPCR (Sotiriou et al., supra). To further confirm the specificity of the primers used for SYBR Green QPCR and the amplicons, standard reverse-transcription PCR was performed with the same primers as the SYBR Green QPCR on a panel of samples that expressed TMPRSS2:ERGα. Similar sized products were obtained and sequencing of cloned products confirmed the presence of TMPRSS2:ERGα. Two cases, PCA16 and PCA17, which expressed high levels of ETV1 or ERG respectively, but showed no evidence of the translocation by QPCR (FIGS. 2, D and E) were identified. RLM-RACE supported these results, as sequencing of the product produced with ETV1 primers in PCA16 revealed no evidence of a fusion transcript and no product could be obtained with ERG primers in PCA17. Similar results were obtained for LNCaP cells, with no evidence of a fusion by RLM-RACE or QPCR, consistent with the exon walking QPCR described above.

Summary of Evidence for TMPRSS2 Fusion Transcripts with ETS Family Members in Prostate Cancer Samples

Results from three different assays for the TMPRSS2:ERG and TMPRSS2:ETV1 fusion transcripts including sequencing of RLM-RACE products, QPCR and sequencing of RT-PCR products are summarized in Table 2. In addition to QPCR for TMPRSS2 fusions being performed in all samples, the existence of these fusions was confirmed using several techniques on selected samples. For example, in PCA1 (prostate cancer sample 1), TMPRSS2:ERGα was identified using sequencing of RLM-RACE products, QPCR and sequencing of RT-PCR products. By QPCR melt curve analysis and gel electrophoresis of QPCR products, PCA4 produced a larger amplicon than expected. Subsequent RLM-RACE analysis confirmed a fusion of the complete exon 1 of TMPRSS2 with the beginning of exon 2 of ERG (TMPRSS2:ERGα) (FIG. 6). Taqman QPCR with the forward primer spanning the TMPRSS2:ERGβ junction confirmed the presence of TMPRSS2:ERGβ only in PCA4 and Taqman QPCR with the forward primer spanning the TMPRSS2:ERGα junction did not produce a product in this specimen (27). Evidence for the TMPRSS2:ERG and TMPRSS2:ETV1 fusions were only found in cases that over-expressed ERG or ETV1 respectively, by QPCR or DNA microarray. These results are in agreement with the exclusive expression observed in the outlier analysis.

Fluorescence In Situ Hybridization (FISH) Confirms TMPRSS2:ETV1 Translocation and ERG Rearrangement

After confirming the existence of the TMPRSS2:ETV1 and TMPRSS2:ERG fusion transcripts, evidence of these rearrangements at the chromosomal level was obtained using interphase fluorescence in situ hybridization (FISH) on formalin fixed paraffin embedded (FFPE) specimens. Two different probe strategies were employed: a two-color, fusion-signal approach to detect TMPRSS2:ETV1 translocations and a two-color, split-signal approach to detect rearrangements of the ERG locus. These probe strategies were validated on the two cases initially used for RLM-RACE, MET26 and MET128 (FIG. 3). Using probes for TMPRSS2 and ETV1, normal peripheral lymphocytes (NPLs) demonstrated a pair of red and a pair of green signals (FIG. 3A). MET26 showed fusion of one pair of signals, indicative of probe overlap (FIG. 3B, yellow arrowhead), consistent with the expression of the TMPRSS2:ETV1 transcript in this case. In addition, consistent low-level amplification of the ETV1 locus was identified, as indicated by the two remaining signals for ETV1 (FIG. 3B, red arrowheads). Similarly, using probes spanning the 5’ and 3’ region of the ERG locus, a pair of yellow signals in NPLs was observed (FIG. 3C). MET28, one pair of probes split into separate green and red signals, indicative of a rearrangement at the ERG locus (FIG. 3D, green and red arrows). This result is consistent with the expression of the TMPRSS2:ERG transcript in this case. Based on these results, the individual FISH analyses described above were performed on serial tissue microarrays containing cores from 13 cases of localized prostate cancer and 16 cases of metastatic prostate cancer (FIG. 3E). As indicated by the matrix, 23 of 29 cases (79.3%) showed evidence of TMPRSS2:ETV1 fusion (7 cases) or ERG rearrangement (16 cases). In addition, 12 of 29 cases (41.4%) showed evidence of low level amplification at the ETV1 locus. Previous reports have identified the genomic location of ETV1, 7p, as one of the most commonly amplified regions in localized and metastatic prostate cancer (Slamon et al., Science 235, 177 (1987)). However, it does not appear that 7p amplification drives ETV1 expression, as ETV1 amplification occurred in 6 cases with ERG rearrangements and our transcript data demonstrates that of 19 samples with high ERG expression and the TMPRSS2:ERG fusion also have high ETV1 expression. Furthermore, when both ETV1 amplification and the TMPRSS2:ETV1 fusion were present by FISH, only the individual ETV1 signal was amplified and not the fused signal. Nevertheless, results from this FISH analysis demonstrate the presence of TMPRSS2:ETV1 and ERG rearrangements at the genomic level consistent with the transcript data described above.

TMPRSS2 is an androgen-regulated gene and fusion with ERG results in androgen regulation of ERG. TMPRSS2 was initially identified as a prostate-specific gene whose expression was increased by androgen in LNCaP cells and also contains androgen responsive elements (AREs) in its promoter (Huang et al., Lancet 361, 1590 (2003); Schwartz et al., Cancer Res 62, 4722 (2002)). Subsequent studies have confirmed high expression in normal and neoplastic prostate tissue and demonstrated that TMPRSS2 is androgen-regulated in androgen-sensitive prostate cell lines (Schwartz et al., Cancer Res 62, 4722 (2002); Ferrando et al., Cancer Cell 1, 75 (2002); Chen et al., Mol Biol Cell 14, 3208 (2003); La Tulippe et al., Cancer Res 62, 4499 (2002)). In addition, while androgen does not increase the expression of TMPRSS2 in the androgen insensitive prostate cancer cell line PC3, stable expression of the androgen receptor in PC3 cells resulted in TMPRSS2 becoming androgen responsive (Schwartz et al., supra; Ferrando et al., supra; Chen et al., supra; La Tulippe et
al., supra). In contrast, microarray studies of LNCaP prostate cell lines treated with androgen have not identified ERG or ETV1 as being androgen-responsive (Jain et al., Cancer Res 64, 3907 (2004)) and examination of their promoter sequences did not reveal consensus AREs (Sotiriou et al., supra). It was contemplated that the TMPRSS2:ERG fusion in DuCaP and VCaP cell lines, which was confirmed by three independent assays in each cell line (Table 2), would result in the androgen regulation of ERG. Using QPCR to assay for ERG expression, it was confirmed that even though ERG was highly expressed in both VCaP and DuCaP cells, treatment with the synthetic androgen R1881 increased the expression of ERG or ETV1 in prostate cancer when respective fusions with TMPRSS2 are present.

Table 1. Cancer Outlier Profile Analysis (COPA). Genes known to undergo causal mutations in cancer that had strong outlier profiles. "X", signifies literature evidence for acquired pathogenomic translocation. "XX" signifies literature evidence for the specific translocation as well as the samples in the specific study that were characterized for that translocation. "Y" signifies consistent with known amplification. "***" signifies ERG and ETV1 outlier profiles in prostate cancer.

<table>
<thead>
<tr>
<th>Rank</th>
<th>%</th>
<th>Score</th>
<th>Study</th>
<th>Cancer</th>
<th>Gene</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95</td>
<td>15.466</td>
<td>Vasselli et al., PNAS USA 100, 6958 (2003)</td>
<td>Renal</td>
<td>PRO1073</td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td>90</td>
<td>12.985</td>
<td>Ross et al., Blood 102, 2951 (2003)</td>
<td>Leukemia</td>
<td>PBN1</td>
<td>XX</td>
</tr>
<tr>
<td>1</td>
<td>95</td>
<td>10.379</td>
<td>Lapointe et al., PNAS USA 101, 811 (Jan 20, 2004)</td>
<td>Prostate</td>
<td>ETV1</td>
<td>**</td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>5.4071</td>
<td>Dhanasekaran et al., Nature 412, 822 (2001)</td>
<td>Prostate</td>
<td>ERG</td>
<td>**</td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>4.3628</td>
<td>Welsh et al., Cancer Res 61, 5974 (2001)</td>
<td>Prostate</td>
<td>ERG</td>
<td>**</td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>4.3425</td>
<td>Zhan et al., Blood 99, 1745 (2002)</td>
<td>Myeloma</td>
<td>CCND1</td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>3.4414</td>
<td>Lapointe et al., supra</td>
<td>Prostate</td>
<td>ERG</td>
<td>**</td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>3.3873</td>
<td>Dhanasekaran et al., Faseb J 19, 243 (2005)</td>
<td>Prostate</td>
<td>ERG</td>
<td>**</td>
</tr>
<tr>
<td>2</td>
<td>90</td>
<td>6.7029</td>
<td>Zhan et al., supra</td>
<td>Prostate</td>
<td>ERG</td>
<td>**</td>
</tr>
<tr>
<td>4</td>
<td>75</td>
<td>2.5728</td>
<td>Sotiriou et al., PNAS USA 100, 10393 (2003)</td>
<td>Breast</td>
<td>ERBB2</td>
<td>Y</td>
</tr>
<tr>
<td>6</td>
<td>90</td>
<td>6.6079</td>
<td>Sotiriou et al., PNAS USA 100, 10393 (2003)</td>
<td>Breast</td>
<td>ERBB2</td>
<td>Y</td>
</tr>
<tr>
<td>9</td>
<td>95</td>
<td>17.698</td>
<td>Ghinasy et al., J Clin Invest 113, 913 (2004)</td>
<td>Prostate</td>
<td>ETV1</td>
<td>**</td>
</tr>
<tr>
<td>9</td>
<td>75</td>
<td>2.2218</td>
<td>Yu et al., J Clin Oncol 22, 2790 (2004)</td>
<td>Prostate</td>
<td>ERG</td>
<td>**</td>
</tr>
</tbody>
</table>

Table 2 shows a summary of TMPRSS2 fusion to ETS family member status in prostate cancer samples and cell lines. For all assays, positive results are indicated by “+” and negative results are indicated by “−”. Blank cells indicate that the specific assay was not performed for that sample. Over-expression of ERG or ETV1 by quantitative PCR (QPCR) is indicated and samples marked with an asterisk indicate the sample was also assessed by cDNA microarray and over-expression was confirmed. In order to detect TMPRSS2:ERG or TMPRSS2:ETV1 gene fusions, selected samples were subjected to RLM-RACE for the over-expressed ETS family member and samples with the TMPRSS2fusion after sequencing are indicated. All samples were assayed for TMPRSS2:ETV1 and
TPRSS2:ERG expression by QPCR. Selected cases were also amplified by standard reverse-transcription PCR (RT-PCR) using the same TPRSS2 fusion primers as for QPCR and amplicons were sequenced. Samples with evidence for TPRSS2:ETV1 or TPRSS2:ERG fusion are indicated in the final column.

**TABLE 2**

<table>
<thead>
<tr>
<th>Case</th>
<th>Sample</th>
<th>QPCR Expression</th>
<th>RLM-RACE sequencing</th>
<th>QPCR TMRSS2:ETV1</th>
<th>QPCR TMRSS2:ERG</th>
<th>RT-PCR sequencing</th>
<th>Family member fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MET26-LN</td>
<td>ETV1*</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>MET26-RP</td>
<td>ETV1*</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>MET28-B</td>
<td>ETV1*</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>MET28-PTLN</td>
<td>ETV1*</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>MET28-41</td>
<td>ETV1*</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>MET16-44</td>
<td>ETV1*</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>MET16-47</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>MET3</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>MET18-23</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>PCA1</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>7</td>
<td>PCA2</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>8</td>
<td>PCA3</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>9</td>
<td>PCA4</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>10</td>
<td>PCA5</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>11</td>
<td>PCA6</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>12</td>
<td>PCA7</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>13</td>
<td>PCA8</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>14</td>
<td>PCA9</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>15</td>
<td>PCA10</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>16</td>
<td>PCA11</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>17</td>
<td>PCA12</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>18</td>
<td>PCA13</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>19</td>
<td>PCA14</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>20</td>
<td>PCA15</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>21</td>
<td>PCA16</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>22</td>
<td>PCA17</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>23</td>
<td>MET30-LN</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>24</td>
<td>MET17-12</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>25</td>
<td>MET20-76</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>26</td>
<td>MET22-61</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>27</td>
<td>MET5-7</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>28</td>
<td>PCA18</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>29</td>
<td>PCA19</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>30</td>
<td>PCA20</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>31</td>
<td>PCA21</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>32</td>
<td>PCA22</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>33</td>
<td>PCA23</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>34</td>
<td>PCA24</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>35</td>
<td>PCA25</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>36</td>
<td>PCA26</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>37</td>
<td>PCA27</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>38</td>
<td>PCA28</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>39</td>
<td>PCA29</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>40</td>
<td>PCA30</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>41</td>
<td>PCA31</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>42</td>
<td>PCA32</td>
<td>ETV1</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

**Cell lines**

- **VCap** ETV1* + + + +
- **DUCaP** ERG - + + +
- **LuCaP** ETV1 - - - -
- **DU145** - - - -
Table 2. Cancer Outlier Profile Analysis (COPA). Genes that are known to undergo causal mutations in cancer that had an outlier profile in the top 10 of a study in Oncomine are shown. “X”, signifies literature evidence for acquired pathogenomic translocation. “XX” signifies literature evidence for the specific translocation as well as that the samples in the specific study were characterized for that translocation. “Y” signifies consistent with known amplification. “***” signifies ERG and ETV1 outlier profiles in prostate cancer.

Table 3

<table>
<thead>
<tr>
<th>Rank</th>
<th>%</th>
<th>Score</th>
<th>Study</th>
<th>Cancer</th>
<th>Reference</th>
<th>Gene</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95</td>
<td>20.056</td>
<td>Valk et al.</td>
<td>Leukemia</td>
<td>Nature 406, 516 (2000)</td>
<td>RUNX1T1</td>
<td>XX</td>
</tr>
<tr>
<td>1</td>
<td>95</td>
<td>15.4462</td>
<td>Vasselli et al.</td>
<td>Renal</td>
<td>PNAS USA 100, 6098 (2003)</td>
<td>PRO1073</td>
<td>X (12)</td>
</tr>
<tr>
<td>1</td>
<td>90</td>
<td>12.9581</td>
<td>Ross et al.</td>
<td>Leukemia</td>
<td>Blood 102, 2951 (2003)</td>
<td>PBX1</td>
<td>XX</td>
</tr>
<tr>
<td>1</td>
<td>95</td>
<td>10.03795</td>
<td>Lapointe et al.</td>
<td>Prostate</td>
<td>PNAS USA 101, 811 (2004)</td>
<td>ETV1</td>
<td>**</td>
</tr>
<tr>
<td>1</td>
<td>90</td>
<td>7.4557</td>
<td>Tian et al.</td>
<td>Myeloma</td>
<td>WHSC1</td>
<td>X (16)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>5.4071</td>
<td>Dhanasekaran et al.</td>
<td>Prostate</td>
<td>Faseb J 19, 243 (2005)</td>
<td>ERG</td>
<td>**</td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>5.2067</td>
<td>Wang et al.</td>
<td>Breast</td>
<td>Lancet 365, 671 (2005)</td>
<td>FOXO3A</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>4.3628</td>
<td>Welsh et al.</td>
<td>Prostate</td>
<td>Cancer Res 61, 5074 (2001)</td>
<td>ERG</td>
<td>**</td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>4.3425</td>
<td>Zhan et al.</td>
<td>Myeloma</td>
<td>Blood 99, 1745 (2002)</td>
<td>CCND1</td>
<td>X (21)</td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>3.724</td>
<td>Cheok et al.</td>
<td>Leukemia</td>
<td>Nat Genet 34, 85 (May, 2005)</td>
<td>PCSK7</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>3.4414</td>
<td>Lapointe et al.</td>
<td>Prostate</td>
<td>PNAS USA 101, 811 (2004)</td>
<td>ERG</td>
<td>**</td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>3.3875</td>
<td>Dhanasekaran et al.</td>
<td>Prostate</td>
<td>Nature 412, 822 (2001)</td>
<td>ERG</td>
<td>**</td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>2.5913</td>
<td>Wigle et al.</td>
<td>Lung</td>
<td>Cancer Res 62, 3005 (2002)</td>
<td>IGH@</td>
<td></td>
</tr>
<tr>
<td>Rank</td>
<td>%</td>
<td>Score</td>
<td>Study</td>
<td>Cancer</td>
<td>Reference</td>
<td>Gene</td>
<td>Evidence</td>
</tr>
<tr>
<td>------</td>
<td>----</td>
<td>-------</td>
<td>-------------</td>
<td>--------</td>
<td>--------------------</td>
<td>-------</td>
<td>----------</td>
</tr>
<tr>
<td>2</td>
<td>95</td>
<td>9.2916</td>
<td>Golub et al.</td>
<td>Leukemia</td>
<td>Science 286, 531 (1999)</td>
<td>TRA@</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>95</td>
<td>9.2916</td>
<td>Golub et al.</td>
<td>Leukemia</td>
<td>Science 286, 531 (1999)</td>
<td>TRD@</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>90</td>
<td>8.2292</td>
<td>Cheok et al.</td>
<td>Leukemia</td>
<td>Nat Genet 34, 85 (May, 2003)</td>
<td>8X2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>95</td>
<td>10.2267</td>
<td>Cheok et al.</td>
<td>Leukemia</td>
<td>Nat Genet 34, 85 (May, 2003)</td>
<td>ARHGP26</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>90</td>
<td>5.9174</td>
<td>Rosenwald et al.</td>
<td>Prostate</td>
<td>Cancer Cell 3, 185 (2003)</td>
<td>REL</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>75</td>
<td>2.6162</td>
<td>Rosenwald et al.</td>
<td>Lymphoma</td>
<td>Cancer Cell 3, 185 (2003)</td>
<td>TCL1A</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>75</td>
<td>2.036</td>
<td>Sotiriou et al.</td>
<td>Breast</td>
<td>PNAS USA 100, 10393 (2003)</td>
<td>RAD51L1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>90</td>
<td>5.4881</td>
<td>Golub et al.</td>
<td>Leukemia</td>
<td>Science 286, 531 (1999)</td>
<td>ICK</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>75</td>
<td>2.0229</td>
<td>Schwartz et al.</td>
<td>Ovarian</td>
<td>Cancer Res 62, 4722 (2002)</td>
<td>IGL@</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>90</td>
<td>17.3733</td>
<td>Ferrando et al.</td>
<td>Leukemia</td>
<td>Cancer Cell 1, 75 (2002)</td>
<td>ZBTB16</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>75</td>
<td>5.7213</td>
<td>La Tulippe et al.</td>
<td>Prostate</td>
<td>Cancer Res 62, 4490 (2002)</td>
<td>NF1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>75</td>
<td>5.2752</td>
<td>Jain et al.</td>
<td>Endocrine</td>
<td>Cancer Res 64, 2907 (2004)</td>
<td>PHD2B</td>
<td></td>
</tr>
</tbody>
</table>
### Table 3-continued

<table>
<thead>
<tr>
<th>Rank</th>
<th>%</th>
<th>Score</th>
<th>Study</th>
<th>Cancer</th>
<th>Reference</th>
<th>Gene</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>95</td>
<td>4.7561</td>
<td>Alon et al.</td>
<td>Colon</td>
<td>Proc Natl Acad Sci USA 96, 6745 (1999)</td>
<td>NONO</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>90</td>
<td>15.3889</td>
<td>Ferrando et al.</td>
<td>Leukemia</td>
<td>Ferrando et al., Cancer Cell 1, 75 (2002)</td>
<td>MN1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>95</td>
<td>10.6036</td>
<td>Segal et al.</td>
<td>Sarcoma</td>
<td>Segal et al., J Clin Oncol 21, 1775 (2003)</td>
<td>KIT</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Oligonucleotide primers used in this study. For all primers, the gene, bases and exons (according to alignment of the reference sequences described in the text with the May 2004 assembly of the human genome using the UCSC Genome Browser) are listed. Forward primers are indicated with “f” and reverse primers with “r”.

### Table 4

<table>
<thead>
<tr>
<th>Gene</th>
<th>Bases</th>
<th>Exon(s)</th>
<th>Primer</th>
<th>Sequence 5' to 3'</th>
<th>SEQ ID NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETV1</td>
<td>193-216</td>
<td>2</td>
<td>Exon 2-3_f</td>
<td>AACAGAGACTGCGTCATGATTCA</td>
<td>1</td>
</tr>
<tr>
<td>ETV1</td>
<td>268-245</td>
<td>3</td>
<td>Exon 2-3_r</td>
<td>CTCTCTGCAAGCCATGTTTCCCTGA</td>
<td>2</td>
</tr>
<tr>
<td>ETV1</td>
<td>248-271</td>
<td>3-4</td>
<td>Exon 3-4_f</td>
<td>AGGAACATGGCTTCAGAGGTC</td>
<td>3</td>
</tr>
<tr>
<td>ETV1</td>
<td>305-280</td>
<td>4</td>
<td>Exon 3-4_r</td>
<td>TCTCTGACAGACTGCTACATTTTGTC</td>
<td>4</td>
</tr>
<tr>
<td>ETV1</td>
<td>269-294</td>
<td>4</td>
<td>Exon 4-5_f</td>
<td>CTCAGGATACCTGACATGATGAGG</td>
<td>5</td>
</tr>
<tr>
<td>ETV1</td>
<td>374-351</td>
<td>5</td>
<td>Exon 4-5_r</td>
<td>CATGGACTGCGGGGTCTTTTCTTG</td>
<td>6</td>
</tr>
<tr>
<td>ETV1</td>
<td>404-429</td>
<td>5</td>
<td>Exon 5-6_f</td>
<td>AACAGCCTTTAAATTCTGCTATGGA</td>
<td>7</td>
</tr>
<tr>
<td>ETV1</td>
<td>492-472</td>
<td>6</td>
<td>Exon 5-6_r</td>
<td>GGGAGGCTTCAACTCCTTGG</td>
<td>8</td>
</tr>
<tr>
<td>ETV1</td>
<td>624-645</td>
<td>6-7</td>
<td>Exon 6-7_f</td>
<td>CTACCACATGGACCACAGATT</td>
<td>9</td>
</tr>
<tr>
<td>ETV1</td>
<td>771-750</td>
<td>7</td>
<td>Exon 6-7_r</td>
<td>CTTAAGGCTTTGTGCTGGAGG</td>
<td>10</td>
</tr>
</tbody>
</table>
EXAMPLE 2

ETV4 Gene Fusions

A. Materials and Methods

ET5 Family Expression in Profiling Studies

To investigate the expression of ETS family members in prostate cancer, two prostate cancer profiling studies were utilized (Lapointe et al., Proc Natl Acad Sci USA 2004; 101:811-6 and Tomlins et al., Science 2005; 310:644-8) present in the Oncomine database (Rhodes et al., Neoplasia 2004; 6:1-6). Genes with an ETS domain were identified by the Interpro filter ‘Ets’ (Interpro ID: IPR000418). Heatmap representations were generated in Oncomine using the ‘median-center per gene’ option, and the color contrast was set to accentuate ERG and ETV1 differential expression.

Samples

Prostate cancer tissues (PCA1-5) were from the radical prostatectomy series at the University of Michigan, which is part of the University of Michigan Prostate Cancer Specialized Program of Research Excellence (S.P.O.R.E.) Tissue Core. All samples were collected with informed consent of the patients and prior institutional review board approval. Total RNA was isolated with Trizol (Invitrogen, Carlsbad, Calif.) according to the manufacturer’s instructions. A commercially available pool of benign prostate tissue total RNA (CPP, Clontech, Mountain View, Calif.) was also used.

Quantitative PCR (QPCR)

QPCR was performed using SYBR Green dye on an Applied Biosystems 7300 Real Time PCR system (Applied Biosystems, Foster City, Calif.) as described (Tomlins et al., supra). The amount of each target gene relative to the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH) for each sample was reported. The relative amount of the target gene was calibrated to the relative amount from the pool of benign prostate tissue (CPP). All oligonucleotide primers were synthesized by Integrated DNA Technologies (Corvalle, Iowa). GAPDH primers were as described (Vandesompele et al., Genome Biol 2002; 3:RESEARCH0034). Primers for exons of ETV4 were as follows (listed 5’ to 3’):

- ETV4_exon4-1: CCGGATGAGGAGGAGGATGA, (SEQ ID NO: 21)
- ETV4_exon4-2: CCGGATGAGGAGGAGGATGA, (SEQ ID NO: 22)
- ETV4_exon4-3: CCGGATGAGGAGGAGGATGA, (SEQ ID NO: 23)
- ETV4_exon4-4: CCGGATGAGGAGGAGGATGA, (SEQ ID NO: 24)
- ETV4_exon4-5: CCGGATGAGGAGGAGGATGA, (SEQ ID NO: 25)
- ETV4_exon4-6: CCGGATGAGGAGGAGGATGA, (SEQ ID NO: 26)
- ETV4_exon4-7: CCGGATGAGGAGGAGGATGA, (SEQ ID NO: 27)
- ETV4_exon4-8: CCGGATGAGGAGGAGGATGA, (SEQ ID NO: 28)

Exons were numbered by the alignment of the RefSeq for ETV4 (NM_001986.1) with the May 2004 freeze of the human genome using the UCSC Genome Browser. For QPCR confirmation of TMPRSS2:ETV4 fusion transcripts, TMPRSS2:ETV4a-f, AATGCTTCTCTTCTTCTCCCCTCCTCTTTGACCTGC (SEQ ID NO: 29) and TMPRSS2:ETV4b-f, AATGCTTCTCTTCTTCTCCCCTCCTCTTTGACCTGC (SEQ ID NO: 30), which detects both TMPRSS2:ETV4a and TMPRSS2:ETV4b transcripts, were used with ETV4_exon4-r, RNA Ligase Mediated Rapid Amplification of cDNA Ends (RLM-RACE) was performed using the GeneRacer RLM-RACE Kit (Invitrogen), according to the manufacturer’s instructions as described (Tomlins et al., supra). To obtain the 5’ end of ETV4, first-strand cDNA from PCA5 was amplified using the GeneRacer 5’ Primer and ETV4_exon4-r or ETV4_exon7-r, GAAGGGGCTGTAGGGGCGACTGT (SEQ ID NO: 31)). Products were cloned and sequenced as described (Tomlins et al., supra). Equivalent 5’ ends of the TMPRSS2:ETV4 transcripts were obtained from both primer pairs.

Fluorescence In Situ Hybridization (FISH)

Formalin-fixed paraffin-embedded (FFPE) tissue sections were used for interphase FISH. Deparaffinized tissue was
treated with 0.2 M HCl for 10 min, 2x SSC for 10 min at 80°C, and digested with Proteinase K (Invitrogen) for 10 min. The tissues and BACs were co-denatured for 5 min at 94°C and hybridized overnight at 37°C. Post-hybridization washing was with 2x SSC with 0.1% Tween-20 for 5 min and fluorescent detection was performed using anti-digoxigenin conjugated to fluorescein (Roche Applied Science, Indianapolis, Ind.) and streptavidin conjugated to Alexa Fluor 594 (Invitrogen). Slides were counterstained and mounted in Pro-Long Gold Antifade Reagent with DAPI (Invitrogen). Slides were examined using a Leica DMRA fluorescence microscope (Leica, Deerfield, Ill.) and imaged with a CCD camera using the CytoVision software system (Applied Imaging, Santa Clara, Calif.).

All BACs were obtained from the BACPAC Resource Center (Oakland, Calif.) and probe locations were verified by hybridization to metaphase spreads of normal peripheral lymphocytes. For detection of TMPRSS2:ETV4 fusion, RP11-35C4 (5' to TMPRSS2) was used with multiple BACs located 3' to ETV4 (distal to ETV4 to proximal: RP11-266124, RP11-224280, and RP11-1001E5). For detection of ETV4 rearrangements, RP11-43643 (5' to ETV4) was used with the multiple BACs 3' to ETV4. Each hybridization, areas of cancerous cells were identified by a pathologist and 100 cells were counted per sample. The reported cell count for TMPRSS2:ETV4 fusions used RP11-242480 and similar results were obtained with both 3' ETV4 BACs. To exclude additional rearrangements in PCA5, FISH was performed with two probes 3' to ETV4 (RP11-266124 and RP11-224280), ERG split signal probes (RP11-95121 and RP11-47617) and TMPRSS2:ETV1 fusion probes (RP11-35C4 and RP11-1241L22). BAC DNA was isolated using a QIAFilter Maxi Prep kit (Qiagen, Valencia, Calif.) and probes were synthesized using digoxigenin- or biotin-nick translation mixtures (Roche Applied Science).

B. Results

The initial CPA screen led to the characterization of TMPRSS2 fusions with ERG or ETV1 (Example 1). It was further contemplated that prostate cancers negative for these gene fusions harbor rearrangements involving other ETS family members. By interrogating the expression of all ETS family members monitored in prostate cancer profiling studies from the Oncomine database (Rhodes et al., supra), marked over-expression of the ETS family member ETV4 was identified in a single prostate cancer case from each of two studies—one profiling grossly dissected tissues (Lapointe et al., supra) (Fig. 7A) and the other profiling laser capture microdissected (LCM) tissues (Fig. 7B). As these cases did not over-express ERG or ETV1, and no benign prostate tissues showed over-expression, it was contemplated that fusion with TMPRSS2 was responsible for the over-expression of ETV4 in these cases. Although ELF3 was also over-expressed in a fraction of prostate cancer cases, in both studies normal prostate tissue samples also showed marked ELF3 over-expression, indicating that a gene fusion driving expression in both benign and cancerous tissue is unlikely. Thus, the ETV4 over-expressing case (designated here as PCA5) was further analyzed.

Total RNA was isolated from PCA5 and exon-walking quantitative PCR was used (QPCR) to confirm the over-expression of ETV4. QPCR demonstrated that exons 3 to exon 2 of ETV4 were markedly over-expressed in this case compared to pooled benign prostate tissue (CUP) (900 fold) and prostate cancers that did not over-express ETV4 and were either TMPRSS2:ERG positive (PCA1-2) or negative (PCA3-4) (Fig. 8A). However, a dramatic decrease (>99%) in the expression of exon 2 of ETV4 relative to distal regions in PCA5 was observed, indicating a possible fusion with TMPRSS2, as observed previously in TMPRSS2:ERG and TMPRSS2:ETV1 positive cases (Tomlins et al., supra).

To identify the 5' end of the ETV4 transcript in PCA5, RNA-ligase mediated rapid amplification of cDNA ends (RLM-RACE) was performed using a reverse primer in exon 7. RLM-RACE revealed two transcripts, each containing 5' ends consisting of sequence located approximately 8 kb upstream of TMPRSS2 fused to sequence from ETV4 (Fig. 8B). Specifically, the 5' end of TMPRSS2:ETV4a has 47 base pairs from this region upstream of TMPRSS2, while the 5' end of TMPRSS2:ETV4b has the same terminal 13 base pairs. These 5' ends of both transcripts were fused to the same contiguous stretch consisting of the 9 base pairs of the intron immediately 5' to exon 3 of ETV4 and the reported reference sequence of exons 3 through the reverse primer in exon 7 of ETV4.

The existence of both transcripts in PCA5 and their absence in CUP and PCA1-4 was confirmed using QPCR. To further exclude the presence of fusion transcripts involving known exons from TMPRSS2, QPCR was performed using a forward primer in exon 1 of TMPRSS2 and the ETV4 exon 4 reverse primer, and as expected, no product was detected in CUP or PCA1-5.

Whether other prostate cancers with ETV4 dysregulation might contain TMPRSS2:ETV4 fusion transcripts structurally more similar to TMPRSS2:ERG and TMPRSS2:ETV1 transcripts (which involve known exons from TMPRSS2) is unknown. The TMPRSS2:ETV4 fusions reported here do not contain the well characterized ARES immediately upstream of TMPRSS2. However, evidence exists for androgen responsive enhancers located upstream of the TMPRSS2 sequences present in the TMPRSS2:ETV4 transcripts described here (Rabbits, Nature 1994; 372:143-9). Nevertheless, the marked over-expression of only ETV4 exons involved in the fusion transcript strongly suggests that the gene fusion is responsible for the dysregulation of ETV4. Together, the structure of the TMPRSS2:ETV4 fusion transcripts supports the conclusion that the regulatory elements upstream of TMPRSS2, rather than transcribed TMPRSS2 sequences, drive the dysregulation of ETS family members.

To confirm the fusion of the genomic loci surrounding TMPRSS2 (21q22) and ETV4 (17q21) as demonstrated by RLM-RACE and QPCR, interphase fluorescence in situ hybridization (FISH) was used. Using probes 5' to TMPRSS2 and 3' to ETV4, fusion of TMPRSS2 and ETV4 loci was observed in 65% of cancerous cells from PCA5 (Fig. 8D). As further confirmation of the rearrangement of ETV4, using probes 3' and 3' to ETV4, 64% of cancerous cells from PCA5 showed split signals. FISH was also performed on PCA5 using two probes 3' to ETV4, ERG split signal probes and TMPRSS2:ETV1 fusion probes to exclude additional rearrangements, with negative results obtained for each hybridization.

Taken together, the results highlight the use of carefully examining outlier profiles in tumor gene expression data, as most analytical methods discount profiles that do not show consistent deregulation (Eisen et al., Proc Natl Acad Sci USA 1998; 95:14863-8; Golub et al., Science 1999; 286:531-7; Tusher et al., Proc Natl Acad Sci USA 2001; 98:5116-21) and would thus fail to identify ETV4 in prostate cancer, which appears rare (2 of 98 cases). Combined with the identification of TMPRSS2:ERG and TMPRSS2:ETV1 fusions, the results presented here show that dysregulation of ETS family members mediated by subversion of ARES or enhancers upstream of TMPRSS2 is a hallmark of prostate tumorigenesis.
EXAMPLE 3

Detection of Gene Fusion RNA

This example describes target capture, amplification and qualitative detection of RNA (IVT) containing the sequences of the four gene fusions in four separate qualitative assays: TMPRSS2:ETV1a, TMPRSS2:ETV1b, TMPRSS2:ERGα and TMPRSS2:ERGβ with APTIMA formulation reagents and HPA detection each spiked with the appropriate target specific oligonucleotides, primers and probes. Table 5 shows sequences of oligonucleotides used in the assay.

<table>
<thead>
<tr>
<th>Gene Fusion</th>
<th>Sequence (5’ to 3’)</th>
<th>SEQ ID NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMPRSS2 exon1/Target Capture</td>
<td>AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTTTCCGAAUUUGCUCCG</td>
<td>59</td>
</tr>
<tr>
<td>TMPRSS2 exon1/Target Capture</td>
<td>AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTTTACGCACTCATAAGGCGAAAACTTTCCGCTGATA</td>
<td>60</td>
</tr>
<tr>
<td>TMPRSS2 exon1/Non-T7</td>
<td>GAGGGCGAGGCGGCGGCGGC</td>
<td>61</td>
</tr>
<tr>
<td>TMPRSS2 exon2/Non-T7</td>
<td>CCAATTCATTCACTGCTGTCAATGATACAGC</td>
<td>62</td>
</tr>
<tr>
<td>ETV1a/b exon4/T7</td>
<td>AATTTTAAATACGACTCACTAATAAGGGGAAACTTTCCGCTGATA</td>
<td>63</td>
</tr>
<tr>
<td>ERGβ exon2/T7</td>
<td>AATTTTAAATACGACTCACTAATAAGGGGAAACTTTCCGCTGATA</td>
<td>64</td>
</tr>
<tr>
<td>ERGα exon4/T7</td>
<td>AATTTTAAATACGACTCACTAATAAGGGGAAACTTTCCGCTGATA</td>
<td>65</td>
</tr>
<tr>
<td>TMPRSS2 exon1:ETV1a Junction/ΔE</td>
<td>GCGGCGGCG-CUCAGCAUUAGCG</td>
<td>66</td>
</tr>
<tr>
<td>TMPRSS2 exon2:ETV1b Junction/ΔE</td>
<td>GCUUGGAACGCA-CUCAAGCUACAGC</td>
<td>67</td>
</tr>
<tr>
<td>TMPRSS2 exon1:ERGα Junction/ΔE</td>
<td>GAGGCCGCGCG-GAACGCUUAACAGG</td>
<td>68</td>
</tr>
<tr>
<td>TMPRSS2 exon1:ERGβ Junction/ΔE</td>
<td>GAGCGCGCGCG-GUAAUCCAGGAUCUUU</td>
<td>69</td>
</tr>
</tbody>
</table>

A. Materials and Methods

RNA Target Capture

Lysis buffer contained 15 mM sodium phosphate monobasic monolysate, 15 mM sodium phosphate dibasic anhydrous, 1.0 mM EDTA disodium dihydrate, 1.0 mM EGTA free acid, and 1.10 mM lithium lauryl sulfate, pH 6.7.

Target capture reagent contained 250 mM HEPES, 310 mM lithium hydroxide, 1.88 M lithium chloride, 100 mM EDTA free acid, at pH 6.4, and 250 µg/ml 1 micromagnetic particles SERA-MAG MG-CM Carbonyl Modified (Sera-dyn, Inc., Indianapolis, Ind.) having dT<sub>14</sub> oligomers covalently bound thereto.

Wash solution contained 10 mM HEPES, 6.5 mM sodium hydroxide, 1 mM EDTA, 0.3% (w/v) ethanol, 0.02% (w/v) methyl paraben, 0.01% (w/v) propyl paraben, 150 mM sodium chloride, 0.1% (w/v) lauryl sulfate, sodium (SDS), at pH 7.5.

RNA Amplification & Detection

Amplification reagent was a lyophilized form of a 3.6 mL solution containing 26.7 mM rATP, 5.0 mM rCTP, 33.3 mM rGTP and 5.0 mM rUTP, 125 mM HEPES, 8% (w/v) threo
dolate dihydrate, 1.33 mM dATP, 1.33 mM dCTP, 1.33 mM dGTP and 1.33 mM dTTP, at pH 7.5. The Amplification reagent was reconstituted in 9.7 mL of the amplification reagent constitution solution (see below). Before use, 15 pmol each of primer oligomers was added.

Amplification reagent constitution solution contained 0.4% (w/v) ethanol, 0.10% (w/v) methyl paraben, 0.02% (w/v) propyl paraben, 35 mM KCl, 30.6 mM MgCl<sub>2</sub>, 0.003% phenol red.

Enzyme reagent was a lyophilized form of a 1.45 mL solution containing 20 mM HEPES, 125 mM N-acetyl-L-
cysteine, 0.1 mM EDTA sodium dihydrate, 0.2% (v/v) TRITON X-100 detergent, 0.2 M trehalose dihydrate, 0.90 RTU/mL Moloney murine leukemia virus (MMLV) reverse transcriptase, and 0.20 U/mL T7 RNA polymerase, at pH 7.0.

One unit (RTU) of activity is defined as the synthesis and release of 5.75 fmol cDNA in 15 minutes at 37°C for MMLV reverse transcriptase, and for T7 RNA polymerase, one unit (U) of activity is defined as the production of 5.0 fmol RNA transcript in 20 minutes at 37°C. Enzyme reagent was reconstituted in 3.6 mL of the enzyme reagent constitution solution (see below).

Enzyme reagent constitution solution contained 50 mM HEPES, 1 mM EDTA, 10% (v/v) TRITON X-100, 120 mM potassium chloride, 20% (v/v) glycerol) anhydrous, at pH 7.0.

Hybridization reagent contained 100 mM succinic acid free acid, 2% (w/v) lithium lauryl sulfate, 100 mM lithium hydroxide, 15 mM aldrihol-2, 1.2 M lithium chloride, 20 mM EDTA free acid, 3.0% (v/v) ethanol, at pH 4.7.

Selection reagent contained 600 mM boric acid, 182.5 mM sodium hydroxide, 1% (v/v) TRITON X-100, at pH 8.5.

The detection reagents comprised detect reagent I, which contained 1 mM nitric acid and 32 mM hydrogen peroxide, and detect reagent II, which contained 1.5 M sodium hydroxide.

B. Assay Protocol

Target Capture

1. Prepare samples by making dilutions of IVT stock solution into STEM at indicated copy levels for 400 µL sample per reaction tube.
2. Using the repeat pipettor, add 100 µL of the TCR with the TCO to the appropriate reaction tube.
3. Using the micropipettor, add 400 µL of each sample to the properly labeled.
4. Cover the tubes with the sealing card(s) and shake the rack gently by hand. Do not vortex. Incubate the rack at 62±1°C in a water bath for 30±5 minutes.
5. Remove the rack from the water bath and blot bottom of tubes dry on absorbent material.
6. Ensure the sealing cards are firmly seated. If necessary, replace with new sealing cards and seal tightly.
7. Without removing sealing cards, incubate the rack at room temperature for 30±5 minutes.
8. Place the rack on the TCS magnetic base for 5 to 10 minutes.

9. Prime the dispense station pump lines by pumping APTIMA Wash Solution through the dispense manifold. Pump enough liquid through the system so that there are no air bubbles in the line and all 10 nozzles are delivering a steady stream of liquid.

10. Turn on the vacuum pump and disconnect the aspiration manifold at the first connector between the aspiration manifold and the trap bottle. Ensure that the vacuum gauge reads greater than 25 in. Hg. It may take 15 seconds to achieve this reading. Reconnect the manifold, and ensure the vacuum gauge is between 7 and 12 in. Hg. Leave the vacuum pump on until all target capture steps are completed.

11. Firmly attach the aspiration manifold to the first set of tips. Aspirate all liquid by lowering the tips into the first TTU until the tips come into brief contact with the bottoms of the tubes. Do not hold the tips in contact with the bottoms of the tubes.

12. After the aspiration is complete, eject the tips into their original tip cassette. Repeat the aspiration steps for the remaining TTUs, using a dedicated tip for each specimen.

13. Place the dispense manifold over each TTU and, using the dispense station pump, deliver 1.0 mL of APTIMA Wash Solution into each tube of the TTU.

14. Cover tubes with a sealing card and remove the rack from the TCS. Vortex once on the multi-tube vortex mixer.

15. Place rack on the TCS magnetic base for 5 to 10 minutes.

16. Aspirate all liquid as in steps 13 and 14.

17. After the final aspiration, remove the rack from the TCS base and visually inspect the tubes to ensure that all liquid has been aspirated. If any liquid is visible, place the rack back onto the TCS base for 2 minutes, and repeat the aspiration for that TTU using the same tips used previously for each specimen.

Primer Annealing and Amplification

1. Using the repeat pipettor, add 75 µL of the reconstituted Amplification Reagent containing the analyte specific primers to each reaction tube. All reaction mixtures in the rack should now be red in color.

2. Using the repeat pipettor, add 200 µL of Oil Reagent.

3. Cover the tubes with a sealing card and vortex on the multi-tube vortex mixer.

4. Incubate the rack in a water bath at 62°±1°C for 10±5 minutes.

5. Transfer the rack into a water bath at 42°±1°C for 5±2 minutes.

6. With the rack in the water bath, carefully remove the sealing card and, using the repeat pipettor, add 25 µL of the reconstituted Enzyme Reagent to each of the reaction mixtures. All reactions should now be orange in color.

7. Immediately cover the tubes with a fresh sealing card, remove from the water bath, and mix the reactions by gently shaking the rack by hand.

8. Incubate the rack at 42°±1°C for 60±15 minutes.

Hybridization

1. Remove the rack from the pre-amplification water bath and transfer to the post-amplification area. Add 100 µL of the reconstituted Probe Reagent with analyte specific probe, using the repeat pipettor. All reaction mixtures should now be yellow in color.

2. Cover tubes with a sealing card and vortex for 5 seconds on the multi-tube vortex mixer.

3. Incubate the rack in a 62°±1°C water bath for 20±5 minutes.

4. Remove the rack from the water bath and incubate at room temperature for 5±1 minutes.

Selection

1. Using the repeat pipettor, add 250 µL of Selection Reagent to each tube. All reactions should now be red in color.

2. Cover tubes with a sealing card, vortex for 10 seconds or until the color is uniform, and incubate the rack in a water bath at 62°±1°C for 10±1 minutes.

3. Remove the rack from the water bath. Incubate the rack at room temperature for 15±3 minutes.

Reading the TTUs

1. Ensure that there are sufficient volumes of Auto Detection Regants I and II to complete the tests.

2. Prepare the LEADER Luminometer by placing one empty TTU in cassette position number 1 and perform the WASH protocol.

3. Load the TTUs into the luminometer and run the HC Rev B protocol.

C. Results

The results are shown in Tables 6-9 for 4 assays with each of the TMPRSS2:ERV and TMPRSS2:ETV1 gene fusion IVTs spiked into TCR.

<p>| TABLE 6 |</p>
<table>
<thead>
<tr>
<th>TMPRSS2:ETV1a (copies IVT/reaction)</th>
<th>RLU</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4,945</td>
</tr>
<tr>
<td>0</td>
<td>4,599</td>
</tr>
<tr>
<td>10</td>
<td>2,185,359</td>
</tr>
<tr>
<td>10</td>
<td>2,268,090</td>
</tr>
<tr>
<td>10</td>
<td>2,284,908</td>
</tr>
<tr>
<td>100</td>
<td>2,270,369</td>
</tr>
<tr>
<td>100</td>
<td>2,302,023</td>
</tr>
<tr>
<td>1,000</td>
<td>2,272,735</td>
</tr>
<tr>
<td>1,000</td>
<td>2,279,627</td>
</tr>
<tr>
<td>1,000</td>
<td>2,285,742</td>
</tr>
</tbody>
</table>

<p>| TABLE 7 |</p>
<table>
<thead>
<tr>
<th>TMPRSS2:ETV1b (copies IVT/reaction)</th>
<th>RLU</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7,743</td>
</tr>
<tr>
<td>0</td>
<td>6,622</td>
</tr>
<tr>
<td>0</td>
<td>7,370</td>
</tr>
<tr>
<td>0</td>
<td>6,181</td>
</tr>
<tr>
<td>0</td>
<td>7,409</td>
</tr>
<tr>
<td>10</td>
<td>7,712</td>
</tr>
<tr>
<td>10</td>
<td>7,178</td>
</tr>
<tr>
<td>10</td>
<td>7,302</td>
</tr>
<tr>
<td>10</td>
<td>8,430</td>
</tr>
<tr>
<td>10</td>
<td>8,331</td>
</tr>
<tr>
<td>100</td>
<td>774,792</td>
</tr>
<tr>
<td>100</td>
<td>285,712</td>
</tr>
<tr>
<td>100</td>
<td>3,261,878</td>
</tr>
<tr>
<td>100</td>
<td>1,349,368</td>
</tr>
<tr>
<td>100</td>
<td>2,757,334</td>
</tr>
<tr>
<td>1,000</td>
<td>3,647,502</td>
</tr>
<tr>
<td>1,000</td>
<td>3,790,087</td>
</tr>
<tr>
<td>1,000</td>
<td>3,813,812</td>
</tr>
<tr>
<td>1,000</td>
<td>3,753,743</td>
</tr>
<tr>
<td>1,000</td>
<td>3,667,242</td>
</tr>
</tbody>
</table>
TABLE 8

<table>
<thead>
<tr>
<th>TMPRSS2:ERG (copies)</th>
<th>RLU</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7,638</td>
</tr>
<tr>
<td>1</td>
<td>7,565</td>
</tr>
<tr>
<td>10</td>
<td>2,043,379</td>
</tr>
<tr>
<td>10</td>
<td>387,408</td>
</tr>
<tr>
<td>10</td>
<td>978,457</td>
</tr>
<tr>
<td>100</td>
<td>2,332,764</td>
</tr>
<tr>
<td>100</td>
<td>2,445,544</td>
</tr>
<tr>
<td>100</td>
<td>2,530,239</td>
</tr>
</tbody>
</table>

TABLE 9

<table>
<thead>
<tr>
<th>TMPRSS2:ERGb (copies)</th>
<th>RLU</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5,978</td>
</tr>
<tr>
<td>10</td>
<td>6,264</td>
</tr>
<tr>
<td>10</td>
<td>2,700,069</td>
</tr>
<tr>
<td>10</td>
<td>2,768,541</td>
</tr>
<tr>
<td>100</td>
<td>2,883,091</td>
</tr>
<tr>
<td>100</td>
<td>2,779,233</td>
</tr>
<tr>
<td>100</td>
<td>2,857,247</td>
</tr>
<tr>
<td>100</td>
<td>2,957,914</td>
</tr>
</tbody>
</table>

EXAMPLE 5

TMPRSS2:ERG Fusion Associated Deletions

This example describes the presence of common deletions located between ERG and TMPRSS2 on chromosome 21q22.2-3 associated with the TMPRSS2:ERG fusion. Associations between disease progression and clinical outcome were examined using a wide range of human PCA samples, 6 cell lines, and 13 xenografts.

A. Materials and Methods

Clinical Samples

Prostate samples used for this study were collected under an IRB approved protocol. All clinically localized PCA samples were characterized by one pathologist and assigned a Gleason score to eliminate inter-observer differences in pathology reporting. Clinically localized PCA samples were collected as part of an ongoing research protocol at the University of Ulm. The hormone refractory samples were taken from the Rapid Autopsy Program of the University of Michigan.

The FISH experiments were conducted on two PCA outcome arrays, which were composed of 897 tissue cores (histospots) from 214 patients. A summary of the patient demographics is presented in Table 10. All patients had undergone radical prostatectomy with pelvic lymphadenectomy at the University of Ulm (Ulm, Germany) between 1989 and 2001. Pre-operative PSA ranged between 1 and 314 ng/mL (mean 36 ng/mL). Mean and maximum follow-up was 3.4 and 8.4 yrs, respectively.

Cell Lines and Xenografts

Androgen independent (PC-3, DU-145, HPV10, and 22Rv1) and androgen sensitive (LNCaP) PCA cell lines were purchased from the American Type Culture Collection (Manassas, Va.) and maintained in their defined medium. HPV10 was derived from cells from a high-grade PCA (Gleason score 4+4=8), which were transformed by transfection with HPV18 DNA (18). 22Rv1 is a human PCA epithelial cell line derived from a xenograft that was serially propagated in mice after castration-induced regression and relapse of the parental, androgen-dependent CWR22 xenograft. The VCAP cell line was from a vertebral metastatic lesion as part of the Rapid Autopsy program at the University of Michigan.

LuCaP 23.1, 35, 73, 77, 81, 86.2, 92.1, and 105 were derived from patients with androgen independent hormone-refractory disease PCA. LuCaP 49 and 115 are from patients with androgen dependent PCA. LuCaP 58 is derived from an untreated patient with clinically advanced metastatic disease and LuCaP 96 was established from a prostate derived tumor growing in a patient with hormone refractory PCA. LuCaP 49 (established from an omental mass) and LuCaP 93 are hormone-insensitive (androgen receptor [AR]-negative) small cell PCAs. These two xenografts demonstrate a neuroendocrine phenotype. LuCaP 23.1 is maintained in SCID mice, and other xenografts are maintained by implanting tumors in male BALB/c nu/nu mice.

Determining TMPRSS2:ERG Fusion Status using Interphase FISH

The FISH analysis for the translocation of TMPRSS2:ERG is described above and previously (Tomlins et al., Science 310:644-8 (2005)). This break apart assay is presented in FIGS. 11 and 14. For analyzing the ERG rearrangement on chromosome 21q22.2, a break apart probe system was...
applied, consisting of the Biotin-14-dCTP labeled BAC clone RP11-24A11 (eventually conjugated to produce a red signal) and the Digoxigenin-dUTP labeled BAC clone RP11-137L13 (eventually conjugated to produce a green signal), spanning the neighboring centromeric and telomeric region of the ER locus, respectively. All BAC clones were obtained from the BACPAC Resource Center, Children's Hospital Oakland Research Institute (CHORI), Oakland, Calif.

Using this break apart probe system, a nucleus without ERG rearrangement exhibits two pairs of juxtaposed red and green signals. Juxtaposed red-green signals form a yellow fusion signal. A nucleus with an ERG rearrangement shows break apart of one juxtaposed red-green signal pair resulting in a single red and green signal for the translocated allele and a combined yellow signal for the non-translocated allele in each cell. Prior to tissue analysis, the integrity and purity of all probes were verified by hybridization to normal peripheral lymphocyte metaphase spreads. Tissue hybridization, washing, and fluorescence detection were performed as described previously (Garraway, et al., Nature 436:117-22 (2005); Rubin, et al., Cancer Res. 64:3814-22 (2004)). At least one TMA core could be evaluated in 59% PCA cases from two TMAs. The technical difficulties with this assay included the absence of diagnostic material to evaluate, weak probe signals, and overlapping cells preventing an accurate diagnosis. The remainder of the analysis focused on the 118 cases of clinically localized PCA that could be evaluated. 15 cases had corresponding hormone naïve metastatic lymph node samples that could also be evaluated.

The samples were analyzed under a 100x oil immersion objective using an Olympus BX-51 fluorescence microscope equipped with appropriate filters, a CCD (charge-coupled device) camera and the CytoVision FISH imaging and capturing software (Applied Imaging, San Jose, Calif.). Evaluation of the tests was independently performed by two pathologists both with experience in analyzing interphase FISH experiments. For each case, it was attempted to score at least 100 nuclei per case. If significant differences between the results of both pathologists were found, the case was refereed by a third pathologist.

Oligonucleotide SNP Array Analysis

Although SNP arrays were intended for genotyping alleles, the SNP array data can provide information on Loss-of-Heterozygosity (Lieberfarb, et al., Cancer Res 63:4781-5 (2003); Lin, et al., Bioinformatics 20:1233-40 (2004)) and detection of copy number alterations (Zhoa, et al., Cancer Cell 3:483-95 (2003)). Using SNP array analysis, it was possible to identify and validate amplified genes in various cancers including melanoma (MITF) (Garraway, et al., Nature 436:117-22 (2005)) and PCA (PDP52) (Rubin, et al., Cancer Res. 64:3814-22 (2004)).

SNP detection on the 100K array began with a reduction in genome representation. Two aliquots of 250 ng of genomic DNA were digested separately with Xbal HindIII. The digested fragments were independently ligated to an oligonucleotide linker. The resulting products were amplified using a single PCR primer under conditions in which 200-2000 bp PCR fragments were amplified. These fragments represent a sub-fraction of the genome. The SNPs tiled on the arrays have been pre-selected as they lie within these Xbal and HindIII fragments and have been validated as robustly detected on the arrays. The derived amplified pools of DNA were then labeled, fragmented further and hybridized to separate HindIII and Xbal oligonucleotide SNP arrays.

Arrays were scanned with a GeneChip Scanner 3000. Genotyping calls and signal quantification were obtained with GeneChip Operating System 1.1.1 and Affymetrix Genotyping Tools 2.0 software. Only arrays with genotyping call rates exceeding 90% were analyzed further. Raw data files were pre-processed and visualized in dChipSNP Lin, et al., Bioinformatics 20:1233-40 (2004)). In particular, pre-processing included array data normalization to a baseline array using a set of invariant probes and subsequent processing to obtain single intensity values for each SNP on each sample using a model based (PM/MM) method (Li, et al., Proc Natl Acad Sci USA 98:31-6 (2001)). Quantitative PCR for TMPRSS2:ERG and TMPRSS2:ETV1 Fusion Transcripts

QPCR was performed using SYBR Green dye (Qigen) on a DNA engine Opticon 2 machine from MJ Research. Total RNA was reverse transcribed into cDNA using TAQMAN reverse transcription reagents (Applied Biosystems) in the presence of random Hexamers. All QPCR reactions were performed with SYBR Green Master Mix (Qigen). All Oligonucleotide primers were designed at Integrated DNA Technologies. Primers that were described by Tomlin et al. (Science 310:644-8 (2005)) and are specific for the fusion were utilized:

TMPRSS2:ERG f.: TACGGCGGCTAGCTGGGAG, (SEQ ID NO: 55)
TMPRSS2:ERG x.: GATAGGCACTCAACAAACGACTGG, (SEQ ID NO: 56)
TMPRSS2:ETV1 f.: CGCGGACTGACGAGGGC, (SEQ ID NO: 57)
TMPRSS2:ETV1 x.: CAGGCCATGAGAAACCAAACTT, (SEQ ID NO: 58)

GAPDH primers were previously described (Vandesompele, et al., Genome Biol 3: RESEARCH0034 (2002)). 10 µMol of forward and reverse primers were used. Procedures were performed according to the manufacturer’s recommended thermocycling conditions. Threshold levels were set during the exponential phase of the QPCR reaction using Opticon Monitor analysis software version 2.02. The amount of each target gene relative to the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH) for each sample was determined using the comparative threshold cycle (CT) method (Applied Biosystems User Bulletin #2). All reactions were subjected to melt curve analysis and products from selected experiments were resolved by electrophoreses on 2% agarose gel.

Statistics

The clinical and pathology parameters were explored for associations with rearrangement status and with the presence of the deletion. Chi-squared test and Fisher exact test were used appropriately. Kaplan-Meier analysis was used to generate prostate-specific antigen recurrence free survival curves of the pathology and the genomic alteration parameters. Log-rank test was used to evaluate statistical significance of associations. Patients with prior neo-adjuvant hormone ablation therapy were excluded. All statistics were performed using SPSS 13.0 for Windows (SPSS Inc., Chicago, Ill.) with a significance level of 0.05.

B. Results

Detection of Deletions on Chromosome 21 Associated with the TMPRSS2:ERG Gene Rearrangement

In order to characterize the frequency of the TMPRSS2:ERG rearrangement in PCA, a modified FISH assay from the assay described by Tomlins, et al. (Science 310:644-8 (2005)) was utilized. The original FISH assay used two probes located on ERG at the centromeric 3' and telomeric 5' ends. The new
observed that in cases with the TMPRSS2:ERG rearrangement with deletion, the deletion was observed in all of the TMA cores from the same patient in 97.9% (59 of 60) of the cases.

### TABLE 10

<table>
<thead>
<tr>
<th></th>
<th>Count</th>
<th>Column N %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤median</td>
<td>55</td>
<td>50,0%</td>
</tr>
<tr>
<td>&gt;median</td>
<td>55</td>
<td>50,0%</td>
</tr>
<tr>
<td><strong>Preoperative PSA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤4.0 ng/ml</td>
<td>6</td>
<td>8,2%</td>
</tr>
<tr>
<td>&gt;4.0 and ≤10 ng/ml</td>
<td>13</td>
<td>17,8%</td>
</tr>
<tr>
<td>&gt;10 ng/ml</td>
<td>54</td>
<td>74,0%</td>
</tr>
<tr>
<td><strong>Gleason Score Sum</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤7</td>
<td>7</td>
<td>6,0%</td>
</tr>
<tr>
<td>8–10</td>
<td>51</td>
<td>63,6%</td>
</tr>
<tr>
<td>&gt;10</td>
<td>59</td>
<td>50,4%</td>
</tr>
<tr>
<td><strong>Nuclear Grade</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>38</td>
<td>35,5%</td>
</tr>
<tr>
<td>3</td>
<td>69</td>
<td>64,5%</td>
</tr>
<tr>
<td><strong>Pathology Stage (PT)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT2</td>
<td>26</td>
<td>22,2%</td>
</tr>
<tr>
<td>PT3a</td>
<td>34</td>
<td>29,1%</td>
</tr>
<tr>
<td>PT3b</td>
<td>57</td>
<td>48,7%</td>
</tr>
<tr>
<td><strong>Surgical Margins status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>78</td>
<td>72,2%</td>
</tr>
<tr>
<td>Positive</td>
<td>34</td>
<td>31,2%</td>
</tr>
<tr>
<td><strong>Lymph Node Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>52</td>
<td>44,1%</td>
</tr>
<tr>
<td>N1</td>
<td>56</td>
<td>47,5%</td>
</tr>
<tr>
<td>N2</td>
<td>10</td>
<td>8,5%</td>
</tr>
<tr>
<td><strong>PSA Recurrence</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>34</td>
<td>48,6%</td>
</tr>
<tr>
<td>Yes</td>
<td>36</td>
<td>51,4%</td>
</tr>
</tbody>
</table>

*Not all data points were available for all 118 cases*

The TMPRSS2:ERG rearrangement was identified in 49.2% of the primary PCA samples and 41.2% in the hormone naïve metastatic LN samples (FIG. 13A). Deletion of the telomeric probe (green signal) (FIG. 1C-D) was observed in 60.3% (56 of) the primary PCA samples and 42.9% (5) of the hormone naïve lymph node tumors with TMPRSS2:ERG rearrangement.

In the 15 cases where there was matched primary and hormone naïve lymph node tumors, there was 100% concordance for TMPRSS2:ERG rearrangement status with 47.7% (7 of 15) of the pairs demonstrating the rearrangement. Deletion of the telomeric (green signal) probe was concordantly seen in 42.9% (3 of 7) of the pairs.

**TMRPSS2:ERG Rearrangement Status and Prostate Cancer Progression**

The associations between rearrangement status and clinical and pathological parameters were observed (FIG. 13). TMPRSS2:ERG rearrangement with deletion was observed in a higher percentage of PCA cases with advanced tumor stage (PT) (p<0.03) (FIG. 13B), and the presence of metastatic disease to regional pelvic lymph nodes (pN, versus pN<sub>−</sub>) (p<0.02). Associations between TMPRSS2:ERG rearrangement with and without deletion and clinical outcome as determined by prostate specific antigen (PSA) biochemical failure for 70 patients where follow up data was available were also assessed. Gleason grade, tumor stage, nuclear grade and lymph node status were good predictors of PSA biochemical failure (all p-values <0.0005). A trend was observed at the univariate level suggesting a PSA recurrence free survival advantage in TMPRSS2:ERG rearranged PCA cases without deletion as determined by the FISH assay.

**EXAMPLE 6**

**TMPRSS2:ERG Gene Fusion Associated with Lethal Prostate Cancer**

In previous studies, the gene fusions of the 5′-untranslated region of TMPRSS2 (21 q22.3) with the ETS transcription
79 factor family members, either ERG (21 q22.2), ETV1 (7p21.2) (Tomlins, et al., Science 310:644-8 (2005)), or ETV4 (Tomlins, et al., Cancer Res. 66(7):3396-400 (2006)) provide a mechanism for the over expression of the ETS genes in the majority of prostate cancers. Furthermore, the fusion of an androgen regulated gene, TMPRSS2, and an oncogene suggests that disease progression may vary based on these molecular subtypes. The most common mechanism for gene fusion is loss of about 2.8 meagabases of genomic DNA between TMPRSS2 and ERG (FIGS. 17A and B). This example describes the risk of metastases or prostate cancer specific death based on the presence of the common TMPRSS2:ERG gene fusion.

A. Methods

The study population comprises men with early prostate cancer (T1a-b, Nx, M0) diagnosed at the Orebro University Hospital, Sweden, between 1977 and 1991 by transurethral resection of the prostate (TURP) or transvesical adenoma enucleation for symptomatic benign prostate hyperplasia as described by Andreén et al. (J. Urol. 175(4):1337-40 (2006)). Baseline evaluation at diagnosis included physical examination, chest radiography, bone scan and skeletal radiography (if needed). Nodal staging was not carried out. Because this evaluation provided no evidence for distant metastases, patients were followed expectantly and received clinical exams, laboratory tests and bone scans every 6 months during the first 2 years after diagnosis and subsequently at 12-month intervals. Patients, who developed metastases, as determined by bone scan, were treated with androgen deprivation therapy if they exhibited symptoms.

The cause of death in the cohort was determined by review of medical records by the study investigators. A validation study regarding cause of death compared to the Swedish Death Register showed greater than 90% concordance, with no systematic under- or over-reporting of any cause of death (Johansson, et al., Lancet 18642;799-805 (1989)). Follow-up of the cohort with respect to mortality was 100% and no patients were lost to follow-up through October 2005. The study endpoint was defined as development of distant metastases or prostate cancer specific death (median follow-up time 9.1 years, maximum 27 years).

All TURP samples were reviewed by one pathologist to confirm a diagnosis of prostate cancer, determine the Gleason score and nuclear grade, and estimate the tumor burden as previously described (J. Urol. 175(4):1337-40 (2006)). A tissue microarray was assembled using a manual arrayer (Rubin, et al., Cancer Epidemiol Biomarkers Prev. 14(6):1424-32 (2005)). The frequency of the TMPRSS2:ERG rearrangement in prostate cancer was assessed using a modified florescence in situ hybridization (FISH) assay from the assay originally described by Tomlins et al. (Science 310:644-8 (2005)). The new assay moved the 5' probe approximately 600 kb in a telomeric direction. At least one TMA core could be evaluated in 92 of the prostate cancer cases.

B. Results

In this population-based cohort of men diagnosed with localized cancer, the frequency of TMPRSS2:ERG fusion was 15.2% (16/105) (FIGS. 17A and B). TMPRSS2:ERG fusion positive tumors were more likely to have a higher Gleason score (two-sided P=0.014) (Table 11). To assess the relation of fusion status and lethal prostate cancer, cumulative incidence regression was used. A significant association between the presence of the TMPRSS2:ERG gene fusion and metastases or disease specific death (FIG. 17C) with a cumulative incidence ratio (CIR) of 3.6 (P=0.004, 95% confidence interval [CI]=1.5 to 8.9) was observed. When adjusting for Gleason Score, the CIR was 2.4 (P=0.07 and 95% CI=0.9 to 6.1). The present invention is not limited to a particular mechanism. Indeed, an understanding of the mechanism is not necessary to practice the present invention. Nonetheless, it is contemplated that, based on the homogeneity of the TMPRSS2:ERG gene fusion in cells in a given tumor and its presence only in invasive prostate cancers (compared to Prostatic Intraepithelial Neoplasia), it is contemplated that this is an early event, which might, in part, contribute to the biology behind the phenotype of the Gleason patterns.

TABLE 11

| Prognostic Factors for a Cohort of Men Expectantly Managed for Localized Prostate Cancer Stratified by the TMPRSS2:ERG Gene Fusion Status |
|--------------------------------------|----------------|----------------|
| Variable                             | Negative       | Positive       |
| No. of patients                      | 78             | 14             |
| Age at diagnosis, y                  | 73 (60 to 103) | 73 (58 to 90)  |
| Gleason Score**                      | 683            |                |
| Gleason Score < 7                    | 48 (61.5%)     | 3 (21.4%)      |
| Gleason Score > 7                    | 20 (25.6%)     | 6 (42.0%)      |
| Pathologic Stage                     | 10 (12.8%)     | 5 (35.7%)      |
| pT1a                                 | 28 (37.4%)     | 2 (14.3%)      |
| pT1b                                 | 90 (64.1%)     | 12 (85.7%)     |
| Nuclear grade***                    |                |                |
| Status***                            | 53 (67.9%)     | 7 (53.8%)      |
| Survived 12 years without metastases or cancer death | 18 (23.1%) | 4 (26.6%) |
| Death due to other causes            | 10 (9.0%)      | 2 (15.4%)      |
| Distant metastases or death due to prostate Cancer | 13 (16.7%) | 7 (50.0%) |

*Clinical parameters of subjects having the TMPRSS2:ERG fusion and of subjects not having the TMPRSS2:ERG fusion were compared by use of 1 tests or chi-square tests for continuous variable and categorical variables, respectively.
**Gleason Score is obtained by summing the major and minor Gleason patterns.
***For one case nuclear grade was not assessed.
****Individuals who lived less than 12 years and have not developed metastases or died of prostate cancer as of October 2005 are classified as long-term survivors. Individuals who lived less than 12 years and did not develop metastases are classified as short-term survivors.

EXAMPLE 7

Detection of TMPRSS2:ETS Fusions in the Urine of Patients with Prostate Cancer

A. Materials and Methods

Urine Collection, RNA Isolation and Amplification

Urine samples were obtained from patients following a digital rectal exam before either needle biopsy or radical prostatectomy. Urine was voided into urine collection cups containing DNA/RNA preservative (Sierra Diagnostics). For isolation of RNA, a minimum of 30 ml of urine were centrifuged at 400 rpm for 15 min at 4°C. RNA later (Ambion) was added to the urine sediments and stored at -20°C until RNA isolation. Total RNA was isolated using a Qiagen RNeasy Micro kit according to the manufacturer’s instructions. Total RNA was amplified using an OmniPlex Whole Transcriptome Amplification (WTA) kit (Rubicon Genomics) according to the manufacturer’s instructions (Tomlins et al., Neoplasia 8:153 [2006]). Twenty five nanograms of total RNA were used for WTA library synthesis and the cDNA library
was subjected to one round of WTA PCR amplification. Amplified product was purified using a QIAquick PCR Purification kit (Qiagen). For cell line proof of concept experiments, the indicated number of VCaP or LNCaP cells was spiked into 1 ml of sterile urine and the samples were processed as for voided urine.

Quantitative PCR

Quantitative PCR (QPCR) was used to detect ERG, ET1, and TMPRSS2:ERG transcripts from WTA amplified cDNA essentially as described (Tomlins et al., Neoplasia 8:153 [2006], Tomlins et al., Science 310:644 [2005], Example 1 above). For each QPCR reaction, 10 ng of WTA amplified cDNA was used as template. Reactions for ERG, ET1, PSA and GAPDH used 2× Power SYBR Green Master Mix (Applied Biosystems) and 25 ng of both the forward and reverse primers. Reactions for TMPRSS2:ERG used 2× Taqman Universal PCR Master Mix and a final concentration of 900 nM forward and reverse primers, and 250 nM probe. For the Taqman assay, samples with Ct values greater than 38 cycles were considered to show no amplification. For all samples, the amount of ERG and ET1 were normalized to the amount of GAPDH. Samples with inadequate amplification of PSA, indicating poor recovery of prostate cells in the urine, were excluded from further analysis. ERG (exon5 – 6 forward) and ET1 (exon6 – 7F) GAPDH², and PSAα primers were as described. The Taqman primers and probe (MGB labeled) specific for TMPRSS2:ERG are as follows:

<table>
<thead>
<tr>
<th>Sample source</th>
<th>Diagnosis</th>
<th>Biopsy Gleason Major</th>
<th>Biopsy Gleason Minor</th>
<th>Biopsy Gleason Score</th>
<th>Pre-Biopsy PSA (ng/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Biopsy</td>
<td>Benign</td>
<td>4.7</td>
<td>8.3</td>
<td>6.7</td>
<td>41</td>
</tr>
<tr>
<td>Pre-Biopsy</td>
<td>Benign</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
</tr>
<tr>
<td>Pre-Biopsy</td>
<td>Benign</td>
<td>4.9</td>
<td>6.7</td>
<td>2.8</td>
<td>3.9</td>
</tr>
<tr>
<td>Pre-Biopsy</td>
<td>Benign</td>
<td>4.8</td>
<td>4.8</td>
<td>3.6</td>
<td>16.6</td>
</tr>
<tr>
<td>Pre-Biopsy</td>
<td>Pca</td>
<td>4.5</td>
<td>4.5</td>
<td>4.5</td>
<td>11.8</td>
</tr>
<tr>
<td>Pre-Biopsy</td>
<td>Pca</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
</tr>
<tr>
<td>Pre-Biopsy</td>
<td>Pca</td>
<td>4.5</td>
<td>4.5</td>
<td>4.5</td>
<td>19.3</td>
</tr>
<tr>
<td>Pre-Biopsy</td>
<td>Pca-treate</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
</tr>
<tr>
<td>Pre-RP</td>
<td>Pca</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
</tr>
<tr>
<td>Pre-RP</td>
<td>Pca</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
</tr>
</tbody>
</table>

Fluorescence In Situ Hybridization (FISH)

Four µm thick formalin-fixed paraffin-embedded (FFPE) sections from matched needle biopsies were used for interphase fluorescence in situ hybridization (FISH), processed and hybridized as described previously (Example 2 and Tomlins et al., Cancer Res 66:3396 [2006]). BAC probes to detect ERG rearrangements, RP1-95121 (5' to ERG) and RP1-476D17 (3' to ERG) were prepared as described previously (Tomlins et al., Cancer Res 66:3396 [2006]; Tomlins et al., Science 310:644 [2005]; Examples 1 and 2 above).

B. Results

This example describes a non-invasive method to detect prostate cancer by the presence of TMPRSS2:ETS fusion transcripts in prostate cancer cells shed into the urine after a digital rectal exam. Results are shown in FIG. 33. As a proof of concept, sterile urine spiked with prostate cancer cell lines expressing high levels of ERG and TMPRSS2:ERG (VCaP) or high levels of ET1 (LNCaP) was used. As shown in FIG. 33A, it was possible to detect ERG over-expression exclusively in VCaP at 1,600 cells and ET1 over-expression exclusively in LNCaP at 16,000 cells by quantitative PCR (QPCR).

By correlating the number of spiked VCaP and LNCaP cells to GAPDH Cₜ (threshold cycle) values, it was observed that, in some cases, urine obtained from patients after a digital rectal exam contained insufficient cell numbers to reliably detect ERG or ET1 over-expression. Thus, total RNA collected from the urine of patients suspected of having prostate cancer was amplified using OmniPlex Whole Transcriptome Amplification before QPCR analysis. Using this strategy, a cohort of 16 patients where urine was obtained after a digital rectal exam before a needle biopsy to detect prostate cancer was assessed. Subsequent assessment of needle biopsies demonstrated that this cohort contained 4 patients with benign prostates, 1 with high grade prostatic intraepithelial neoplasia (HGPIN) and 11 with prostate cancer. In addition, a cohort of 3 patients with prostate cancer where urine was collected after a digital rectal exam before radical prostatectomy was assessed.

Cohort characteristics are presented in Table 12. Each urine specimen was from a unique patient and was assigned an ID. The source of the sample (pre biopsy or radical prostatectomy (RP) is indicated. The diagnosis following needle biopsy (including benign, high grade prostatic intraepithelial neoplasia (HGPIN), and prostate cancer (PCa)) is indicated. For patients diagnosed as having prostate cancer following needle biopsy, major Gleason, minor Gleason, and Gleason sum score are indicated. For all patients, pre biopsy PSA (ng/ml) and age are reported, if available.

From the needle biopsy cohort, 5 patients were identified with marked over-expression of ERG, 1 of which was diagnosed by needle biopsy as having HGPIN, while the other 4 were diagnosed as having prostate cancer. From the radical prostatectomy cohort, 1 of 3 patients with prostate cancer were identified as having high ERG expression (FIG. 33B). ET1 over-expression was not detected in any patients from either cohort. To confirm the expression of TMPRSS2:ERG in the samples which over-expressed ERG, a TaqMan primer/probe assay designed to specifically amplify TMPRSS2:ERG was utilized. This assay robustly amplified product from VCaP cells, which express TMPRSS2:ERG (Tomlins et al., Science 310:644 [2005]). In addition, 5 of the 6 urine samples from patients with prostate cancer that over-expressed ERG also expressed TMPRSS2:ERG (Ct values 29.8-38.9), while 0 of the 10 samples from patients without ERG over-expression expressed TMPRSS2:ERG. As one sample over-expressed ERG without expression of TMPRSS2:ERG, it is likely that this sample expresses other isoforms of the fusion transcript, such as TMPRSS2:ERGβ or more recently identified fusion transcripts (Soller et al., Genes Chromosomes Cancer 45:717 [2006]; Yoshimoto et al., Neoplasia 8:465:2006). To confirm that the presence of TMPRSS2:ERG fusion transcripts indicates the presence of TMPRSS2:ERG positive cancerous tissue, fluorescence in situ hybridization (FISH) was performed using probes
designed to detect ERG rearrangements on matched tissue sections from representative cases. Matched tissue was obtained from three patients with detectable TMPRSS2:ERG transcripts in the urine and a diagnosis of cancer, one patient with detectable TMPRSS2:ERG transcripts in the urine and a diagnosis of high grade PIN, and two patients without detectable TMPRSS2:ERG transcripts and a diagnosis of cancer. As shown in FIG. 33B, both patients diagnosed with cancer but without detectable TMPRSS2:ERG transcripts in their urine did not harbor ERG rearrangements in cancerous tissue by FISH. All three patients diagnosed with cancer and with detectable TMPRSS2:ERG transcripts in their urine also showed ERG rearrangements in cancerous tissue by FISH. Finally, the patient with a diagnosis of high grade PIN with detectable TMPRSS2:ERG in their urine did not show ERG rearrangements in high grade PIN tissue. This indicates that this patient may have undiagnosed cancer elsewhere in the prostate, resulting in the presence of detectable TMPRSS2:ERG transcripts in their urine.

EXAMPLE 8

TMPRSS2 and ETS Family Genes Fusions in Prostate Cancer

This study describes a comprehensive analysis of the frequency for the TMPRSS2 and ETS family genes rearrangements in a screening-based cohort of 111 American men surgically treated for clinically localized prostate cancer.

A. Materials and Methods

Study Population, Clinical Data and Prostate Sample Collection:

As a source of clinically localized prostate cancers, a tissue microarray (TMA) containing—cores representing cancer and benign tissue was constructed from 111 men who underwent radical prostatectomy at the University of Michigan as the primary monotherapy (i.e., no adjuvant or neoadjuvant hormonal or radiation therapy). The radical prostatectomy series is part of the University of Michigan Prostate Cancer Specialized Program of Research Excellence (SPORE) Tissue Core. Three cores (0.6 mm in diameter) were taken from each representative tissue block to construct the TMA. The TMA construction protocol has been described (Kononen et al., Nat. Med. 4:844 [1998]; Rubin et al., Am J Surg Pathol 26:312 [2002]). Detailed clinical, pathological, and TMA data were maintained on a secure relational database as previously described (Manley et al., Am J Pathol 159:837 [2001]).

Assessment of TMPRSS2-ETS Gene Fusion Using an Interphase Fluorescence In Situ Hybridization Assay

Four µm thick tissue micro array sections were used for interphase fluorescence in situ hybridization (FISH), processed and hybridized as described previously (Tomlins et al., Science 310:644 [2005]; Tomlins et al., Cancer Res 66:3396 [2006]). Slides were examined using an Axiosoplan ImagingZ1 microscope (Carl Zeiss) and imaged with a CCD camera using the ISIS software system in Metafer image analysis system (Meta Systems, Altluessheim, Germany). FISH signals were scored manually (100x oil immersion) by pathologists in morphologically intact and non-overlapping nuclei and a minimum of 50 cells or the maximum numbers of cancer cells available in three cores from a case were recorded. Cases without 30 evaluable cells were reported as insufficient hybridization. All BACs were obtained from the BACPAC Resource Center (Oakland, Calif.), and probe locations were verified by hybridization to metaphase spreads of normal peripheral lymphocytes. For detection of TMPRSS2, ERG and ETV4 rearrangements we used the following probes: RP11-35C4 (5' to TMPRSS2) and RP11-120C17 (3' to TMPRSS2), RP11-95121 (5' to ERG) and RP11-476D17 (3' to ERG), and RP11-100E5 (5' to ETV4) and RP11-456D4 (3' to ETV4). For detection of TMPRSS2:ETV1 fusion, RP11-35C4 (5' to TMPRSS2) was used with RP11-1241.22 (3' to ETV1). BAC DNA was isolated using a QIAfilter Maxi Prep kit (Qiagen, Valencia, Calif.), and probes were synthesized using digoxigenin- or biotin-nick translation mixtures (Roche Applied Science, Indianapolis, Ind.). The digoxigenin and biotin labeled probes were detected using fluorescein conjugated anti-digoxigenin antibodies (Roche Applied Science) and Alexa 594 conjugated streptavidin (Invitrogen, Carlsbad, Calif.), respectively.

A break apart (TMPRSS2, ERG, ETV4) or fusion (TMPRSS2-ETV1) probe strategy was employed to detect rearrangements at the chromosomal level. Normal signal patterns for TMPRSS2, ERG and ETV4 in DAPI stained nuclei were indicated by two pairs of colocalized green and red signals. For these probes, a rearrangement was confirmed by break apart of one of the two colocalized signals. For TMPRSS2-ETV1 fusion, two pairs of separate red and green were recorded as normal, while one pair of separate and one pair of colocalized signals was recorded as a rearrangement.

B. Results and Discussion

This example describes a comprehensive analysis outlining the signature of TMPRSS2 and ETS transcription factor genes rearrangement in a large screening-based cohort of American men surgically treated for clinically localized prostate cancer. A TMPRSS2 split probe FISH assay approach was used to detect the overall frequency of gene rearrangement in prostate cancer with known ETS family partners ERG, ETV1, ETV4 and other unknown partners, as shown in FIG. 34. It was hypothesized that prostate cancers negative for three known ETS partners (ERG, ETV1 and ETV4) may harbor rearrangements involving other ETS family members. The results demonstrate complex molecular signature of TMPRSS2 and ETS family genes rearrangement in clinically localized prostate cancer (FIGS. 35A and B). Overall TMPRSS2 was rearranged in 65% of evaluable cases, while ERG, ETV1 and ETV4 were rearranged in 55%, 2% and 2% of evaluable cases (FIG. 35A). In 40.5% of cases with TMPRSS2 rearrangement, loss of the 3' probe was observed, consistent with a chromosomal deletion between TMPRSS2 and ERG as a mechanism of gene fusion. These results confirm the high frequency of TMPRSS2:ETS fusions in prostate cancer and confirm previous studies showing that TMPRSS2: ERG are by far the most common type (Tomlins et al., Science 310:644; Perner et al., Cancer Res 66:3396 [2006]; Yoshimoto et al., Neoplasia 8:4665 [2006]; Soller et al., Genes Chromosomes Cancer 45:717 [2006]; Wang et al., Cancer Res 66:8347 [2006] and above examples).

Similar results were observed when the cohort was limited to just those cases where all four probes were evaluable (FIGS. 35A and B). This analysis confirmed that TMPRSS2: ETS rearrangements are mutually exclusive, as no cases showed rearrangements of multiple ETS family members. This analysis also demonstrates that a single TMPRSS2 assay can effectively detect almost all ETS rearrangements, as the 24 cases with ERG, ETV1 or ETV4 rearrangement were detected by the TMPRSS2 assay. In all 9 cases where the 3' ERG probe was deleted, deletion of the 3' TMPRSS2 probe was identified.

Furthermore, two cases were identified with break apart of the TMPRSS2 probes, indicating a rearrangement, without rearrangement of ERG, ETV1 or ETV4 (cases 32 and 36) and cases with TMPRSS2 rearrangement without ERG rear-
arrangement where ETV1 and/or ETV4 could not be evaluated. These cases suggest that TMPRSS2 may be partnering with novel ETS family members or unrelated oncogenes in prostate cancer. Together, these results suggest that a single TMPRSS2 assay can provide diagnostic and prognostic information in prostate cancer.

EXAMPLE 9

PSA Gene Fusions

FISH experiments were used to identify cases that show a split signal by FISH for probes located 5' and 3' to PSA. The 5' and 3' BACs used to detect the PSA split are RP11-510116 and RP11-26P14, respectively. A partner for the PSA gene fusion has not yet been identified. These same probes also pick up a split in the ETS family member SPIB, as it is located very close to PSA.

EXAMPLE 10

FL11 Overexpression

FL11 expression was assayed in different cell samples not harboring a FL11 gene fusion. The expression of 5' and 3' exons of FL11 was measured from a case with high FL11 expression. Results are shown in FIG. 18. No difference in the 5' and 3' transcript abundances was detected. RACE also did not indicate a fusion transcript. FL11 was overexpressed in prostate cancer relative to control samples. Primers for FL11 amplification, as well as TaqMan probes, are shown in FIG. 37.

FISH was also used to identify samples that have split signals for FL11, indicating a rearrangement, but these cases do not have TMPRSS2:FL11 fusion by FISH. BAC probes are shown in Table 13. These cases also have high FL11 expression.

EXAMPLE 11

Tissue Microarrays

Tissue microarrays were used to assay for the presence of gene fusions. TMAs used included prostate cancer progression array, prostate cancer outcome array, urothelial array, prostate cancer screening array, and negative prostate cancer array, and individual prostate cancer cases. The following gene probes were used on tissue microarrays: Tmprss2-ETV1 fusion probes, ERG split probes, TMPRSS2 split probes, ETV1 split probes, ETV4 split probes, and FL1 split probes.

In addition, ERG split probes were used on an outcome array. The results are as follows: negative cases: 30, positive cases: 29, marginal cases: 1. There was a weak association of ERG positive cases with higher Gleason score (≥ 7).

Protein arrays and mass spec were used to identify nuclear interactors for ERG2. The results are shown in FIG. 21.

EXAMPLE 12

Androgen Regulation of Erg Expression

This Example describes the androgen regulation of Erg expression. LNCap (TMPRSS2-ERG−) and VCaP (TMPRSS2-ERG+) cell lines were used. The cells were contacted with varying amounts of R1881 for 48 hrs. Expression of Erg, PSA (+ control) and beta-tubulin (+ control) were assayed.

The results are shown in FIG. 19. ERG expression was found to be androgen dependent in the VCaP, but not the LNCap cells.

EXAMPLE 13

Peptide Antibody and Aqua Probe Generation

FIGS. 22-25 shows sequences (underlined) of ERG1, ETV1, FL1-1, and ETV4 for use in peptide antibody generation and for making aqua probes. Primers are designed by Applied Biosystems for all ETS family members. Expression is monitored in prostate cancer cases, with high expression being an indicator of a possible gene fusion and an indicator for FISH.

EXAMPLE 14

ETV1 in LnCaP Cells

This Example describes an analysis of the transcriptional response to androgen in VCaP and LNCaP. In addition to detecting a number of transcripts differentially expressed in both cell lines were identified, such as PSA, a number of transcripts uniquely dysregulated in VCaP or LNCaP cells were also identified. This analysis identified ETV1 as being exclusively responsive to androgen in LNCaP cells. Combined with the over-expression of ETV1 in LNCaP cells, FISH was used to interrogate the ETV1 loci in LNCaP cells.

A: Materials And Methods

Cell Lines

The prostate cancer cell lines LNCaP (originally derived from a lymph node prostate cancer metastasis) and VCaP (Korenchuk, S. et al., In vivo 15, 163-8 (2001)) (originally derived from a vertebral prostate cancer metastasis) were used for this study. For microarray studies, VCaP and LNCaP cells were grown in charcoal-stripped serum containing media for 24 hours before treatment for 48 hours with 0.1% ethanol or 1 nM of the synthetic androgen methyltrienolone (R1881, NEN Life Science Products, Boston, Mass.) dissolved in ethanol. For quantitative PCR (QPCR) studies, cells were grown in charcoal-stripped serum containing media for 24 hours, preincubated with 0.1% ethanol, Casodex dissolved in acetone (10 uM, bicalutamide, AstraZeneca Pharmaceuticals, Wilmington, Del.) or flutamide dissolved in ethanol (10 uM, Sigma, St. Louis, Mo.). After 2 hours, 0.1% ethanol or 0.5 nM of R1881 was added and the cells were harvested after 48 hours. Total RNA was isolated from all samples with Trizol (Invitrogen, Carlsbad, Calif.) according to the manufacturer’s instructions. RNA integrity was verified by denaturing formaldehyde gel electrophoresis or the Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, Calif.).

Microarray Analysis

The cDNA microarrays used for this study were constructed essentially as described, except the array contains 32,448 features. Protocols for printing and post-processing of arrays are available on the Internet. cDNA microarray analysis was done essentially as described. Briefly, total RNA from control and R1881 treated VCaP and LNCaP cell lines were reverse transcribed and labeled with Cy5 fluorescent dye. Pooled total RNA from control VCaP or LNCaP samples were reverse transcribed and labeled with Cy3 fluorescent dye for all hybridizations from the respective cell lines. The labeled products were then mixed and hybridized to the cDNA arrays. Images were flagged and normalized using the Genepix software package (Axon Instruments Inc., Union City, Calif.). Data were median-centered by arrays and only
genes that had expression values in at least 80% of the samples were used in the analysis.

Quantitative PCR (QPCR)

QPCR was performed using SYBR Green dye on an Applied Biosystems 7300 Real Time PCR system (Applied Biosystems, Foster City, Calif.) as described (Tomlins et al., Cancer Res 66, 3396-400 (2006); Tomlins et al., Science 310, 644-8 (2005)). The amount of each target gene relative to the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH) for each sample was reported. The relative amount of the target gene in each cell line and/or experiment was calibrated to controls. All oligonucleotide primers were synthesized by Integrated DNA Technologies (Coralville, Iowa). GAPDH (Vandesompele et al., Genome Biol 3, RESEARCH0034 (2002)), PSA (Specht et al., Am J Pathol 158, 419-29 (2001)), ERG (Eton 6-6 and Eton 6-6) and ETV1 (Exon 6-7 and Exon 6-7) primers (Tomlins et al., Science 310, 644-8 (2005)) were as described.

Fluorescence In Situ Hybridization (FISH)

Metaphase spreads were prepared from normal peripheral lymphocytes (NPLs) and LNCaP cells using standard techniques. Slides were treated with 2×SSC for 2 min, 100% ethanol for 2 min and 100% ethanol for 2 min before addition of the probe. Slides were coverslipped and incubated at 75°C for 2 min and hybridized overnight at 37°C. Post-hybridization washing was with 2×SSC at 42°C. For 5 min, followed by 3 washes in PBST. Fluorescent detection was performed using anti-digoxigenin conjugated to fluorescein (Roche Applied Science, Indianapolis, Ind.) and streptavidin conjugated to Alexa Fluor 594 (Invitrogen, Carlsbad, Calif.). Slides were counterstained and mounted in ProLong Gold Antifade Reagent with DAPI (Invitrogen). Slides were examined using a Zeiss Axio Imager Z1 fluorescence microscope (Zeiss, Thornwood, N.Y.) and imaged with a CCD camera using ISIS software (Metasystems, Altlussheim, Germany).

All BACs were obtained from the BACPAC Resource Center (Oakland, Calif.) and probe locations were verified by hybridization to metaphase spreads of normal peripheral lymphocytes. For hybridization to the ETV1 region on chromosome 7p, four BACs were used (telomeric to centromeric): RP11-124L22, RP11-313C20, RP11-703A4 and RP11-1149J13. For localization to chromosome 14q, the FISH mapped BAC RP11-483K13, which we also confirmed as hybridizing to 14q using NPLs. BAC DNA was isolated using a QIAfilter Maxi Prep kit (Qiagen, Valencia, Calif.) and probes were synthesized using digoxigenin- or biotin-nick translation mixtures (Roche Applied Science).

B. Results

Results are shown in FIGS. 26-28. FIG. 26 shows the over-expression and androgen regulation of ETV1 in the LNCaP prostate cancer cell line. FIG. 26A shows expression signature of androgen-regulated genes in VCaP and LNCaP prostate cancer cell lines. Heatmap of genes showing induction or repression in either cell line (3,499 features, p<0.05 and fold change ratio—1.5) by 1 nM synthetic androgen R1881 (green) compared to vehicle treatment (gray). Each row represents a gene; each column represents a sample. Yellow and blue cells indicate over- or under-expression, respectively, according to the color scale. Gray cells indicate missing data. Values for each cell line are centered on the corresponding control samples. The locations of PSA, ERG and ETV1 in the heatmap are indicated and their expression is shown in the inset. FIG. 26B shows confirmation of PSA induction by androgen in both VCaP and LNCaP cells by quantitative PCR (QPCR). The relative expression of PSA (normalized to GAPDH) in LNCaP (red) and VCaP (blue) cell lines was determined by QPCR. Cells were treated with vehicle or 1 nM R1881 for 48 hours in the presence or absence of the anti-androgens Casodex or Flutamide as indicated. The relative amount of PSA in each sample was calibrated to the amount in the control sample for each cell line. FIG. 26C shows ETV1 induction by androgen in LNCaP cells. Using the same samples as B, the relative amount of ETV1 was determined by QPCR. FIG. 26D shows that ETV1 is markedly over-expressed in LNCaP cells. The relative expression of PSA and ETV1 and ERG were determined in the 48 hour control samples from each cell line by QPCR. The relative amount of target gene in each sample was calibrated to the average amount of PSA from both cell lines. The fold difference in ETV1 and ERG expression in LNCaP cells is indicated.

FIG. 27 shows rearrangement of ETV1 in LNCaP cells. FIG. 27A shows a schematic of BACs used as probes for fluorescence in situ hybridization (FISH). The location and coordinates at 7p21 (including the ETV1 locus and surrounding BACs) and 14q32 was determined on the May 2004 freeze of the human genome using the UCSC Genome Browser. BACs used in this study are indicated as numbered rectangles. The location of ETV1 and DKG2 are shown with the arrowhead indicating the direction of transcription. FIG. 27B shows that RP11-124L22 and RP11-1149J13 co-localize to chromosome 7 in normal peripheral lymphocytes (NPLs). Localization of RP11-124L22 (BAC #1) and RP11-1149J13 (BAC #4) on metaphase spreads (top panel) or interphase cells (bottom panel) was determined by FISH in NPLs. For all metaphase pictures, signals on chromosome 7 are indicated by arrows, while signals on chromosome 14 are indicated by arrowheads of the corresponding probe color. Higher magnification of informative regions of metaphase spreads are shown in boxes. FIG. 27C shows localization of BAC #1 and BAC #4 on metaphase spreads (top panel) and interphase cells (bottom panel) was determined in the near tetraploid LNCaP cell line. Two co-localized signals on chromosome 7, two red signals on chromosome 7 and two green signals on a different chromosome were observed. FIG. 27D shows signal from RP11-124L22 localizes to chromosome 14 in LNCaP cells. As in C, except RP11-124L22 (BAC #1) was co-hybridized with RP11-483K13 (BAC #5, FISH mapped to chromosome 14q) on LNCaP metaphase spreads. Four red signals from RP11-483K13 localize to chromosomes 14q; two green signals localize to chromosome 7p and two green signals localize to chromosome 14q.

FIG. 28 shows that the entire ETV1 locus is inserted into chromosome 14 in LNCaP cells. FIG. 28A shows a schematic of BACs used in this experiment. FIG. 28B shows localization of RP11-124L22 (BAC #1) and RP11-313C20 (BAC #2) on metaphase spreads (top panel) and interphase cells (bottom panel) was determined by FISH in LNCaP cells. In metaphase spreads, two pairs of co-localized signals were observed on chromosome 7 (yellow arrows) and chromosome 14 (yellow arrowheads).

These results demonstrate that the entire ETV1 locus is translocated from chromosome 7 to chromosome 14. Although the genomic sequence upstream of the insertion on chromosome 14 is unknown, it is likely that this region contains AREs, which drive the high level of ETV1 observed only in LNCaP cells and the androgen responsiveness. These results suggest that LNCaP cells find use as an in vitro model of ETS gene fusions seen in human prostate cancers.

EXAMPLE 15

Knockdown of ETS Family Members in PCA

This Example describes the knockdown of ETS family members in prostate cancer. siRNAs were used to knockdown
expression of ETV1 and ERG in LncAP and VCAP. Quantitative PCR was used to confirm the knockdown. Results are shown in FIGS. 29 and 30. The knockdown did not affect proliferation. Lentivirus expressing shRNA are generated for stable knockdowns.

Microarrays were performed on Agilent 44K Whole Genome arrays to determine which genes were differentially expressed when ERG expression was knocked down in VCaP cells (which have the TMPRSS2:ERG fusion). For this experiment, three conditions were used: knockdown using Dharmacon siRNA for ERG (ERGsi), knockdown of luciferase (control), and untransfected (untrans) VCaP cells. Three hybridizations of ERG/untrans and two of control/untrans were performed. The genes were called as present in all five experiments, had standard deviations less than 0.5 (of the average for both conditions), and showed a fold difference between the ERG and control of <0.75 or >1.5. The ERGdiff field indicates the fold difference between the ERG and control knockdown experiments, so value less than one means the gene is underexpressed in the ERG knockdown (ERG itself ranks 81st in this analysis).

EXAMPLE 16

Transgenic Mice

Transgenic mice that over express gene fusions of the present invention, as well as ETS and androgen responsive genes are generated. FIG. 31 shows viral overexpression systems for use in generating mice. FIG. 32 shows a schematic of genomic insertions in transgenic mice. Such mice find use in research (e.g., mechanistic studies) and drug screening applications.

EXAMPLE 17

Identification of TMPRSS2:ERGa

As described above (Example 1), fusions of TMPRSS2 to ERG were observed. To determine the expressed protein from the TMPRSS2:ERGa gene fusion, PCR was used to amplify the portion of ERG (NM_004449) from the fusion breakpoint at the beginning of exon 4 to the presumed stop codon in exon 11, inserting a 3x Flag tag immediately upstream of the stop codon, from the VCaP prostate cancer cell line. The product was TA cloned into pCR8/GW/TOPO TA (Invitrogen) and bi-directionally sequenced. Sequencing revealed the presence of two distinct isoforms, herein designated as ERG1 (includes exon 6 from ERG isoform 1 (NM_182918, GGGGTACGATCTTATTTATTTACCGAAGC- TAGCAAGAATTACAACTAGGACCCAG; SEQ ID NO:73) and ERG2 (does not include this exon). The product was Gateway cloned into the pLent6/V5-DEST destination vector. This plasmid was transfected directly into PHIXN cells for ERG protein production.

A. Methods

Transfection Assay. Phenx cells were transfected with either ERG2 or the empty vector using Fugene transfection reagent (Roche) as per manufacturer’s instructions. A total of ten 150 mm diameter plates were used for each construct. The cells were harvested 48 h post-transfection and used for immunoprecipitation assay as described below.

Protein Lysis and Immunoprecipitation: Cells were washed in ice cold PBS containing protease inhibitors and lysed by homogenization in TBS containing 1% NP40. The supernatant containing proteins were estimated for their protein content using Bradford’s Protein Assay (Biorad Laboratories, Hercules, Calif.) as per manufacturer’s instructions. Equal amounts of protein (approximately 30 mg in 15 ml buffer) from all samples were used for immunoprecipitation studies. About 200 microtubes of a 50% slurry of EZView Red ANTI-FLAG M2 Affinity Gel (Sigma, St Louis, Mo.) was added to each sample and incubated overnight at 4 C. The immunoprecipitate was washed thrice each with TBS containing 0.1% NP40 and TBS alone. Bound proteins were eluted using FLAG peptide (Sigma, St Louis, Mo.) as per manufacturer’s instruction. The elution was performed three times. Proteins in the eluate were precipitated using 50% TCA (Sigma, St Louis, Mo.). The precipitate was washed thrice with ice cold acetone, redissolved in Laemmli buffer and electrophoresed on 4-20% BIS-TRIS gel (Invitrogen Corporation, Carlsbad, Calif.). The gels were stained with mass spectrometry compatible silver stain (Silver Quest, Invitrogen Corporation, Carlsbad, Calif.). Bands corresponding to ERG2 and the corresponding region in the vector lane were excised into 6 pieces of 1 cm each. Each of the gel pieces were labeled bands 1-6 starting from higher molecular weight region on the gel moving down. Thus Band 1 corresponds to the region containing high molecular weight proteins while band 6 corresponds to region of low molecular weight. Based on its native molecular mass of ERG2 (approximately 55 KDa) would migrate in Bands 4 and 5. ERG2 sequence identification was repeated three times and the data was consolidated from all the experiments.

Protein Identification

The gel bands were collected, destained using the destaining solution provided in the Silver Stain Kit as per manufacturers instruction (Invitrogen Corporation, Carlsbad, Calif.). In gel digestion was performed using Porcine Trypsin (1:50, Promega Corporation, Madison, Wis.) in 1M Ammonium Bicarbonate, pH 9. The digestion was performed for 16 h at 37 C. At the end of 24 h the trypsin activity was stopped using 3% formic acid. The peptides were extracted using 50% Acetonitrile. The peptides were dried and redissolved in 2% Acetonitrile containing 0.1% formic acid and separated by reversed-phase chromatography using a 0.075 mm x 150 mm C18 column attached to a Paradigm HPLC pump (Michrome Bio Resources Inc.). Peptides were eluted using a 45-min gradient from 5 to 95% B (0.1% formic acid/95% acetonitrile), where solvent A was 0.1% formic acid/2% acetonitrile. A Finnigan LTQ mass spectrometer (Thermo Electron Corp.) was used to acquire spectra, the instrument operating in data-dependent mode with dynamic exclusion enabled. The MS/MS spectra on three most abundant peptide ions in full MS scan were obtained. The spectra are searched using the MASCOT search tool against the composite, non-identical NCB1 human reference sequence database. These database search results are validated for peptide assignment accuracy using the PeptideProphet program. This is a mixture model; an expectation maximization evaluation assigning a probability of correct peptide identification based on search result scores and various peptide features including the number of typtic termini. A second program, ProteinProphet, is used to group peptides by protein and combine their probabilities to assign a probability of a correct protein assignment. Discriminatory power increases with the subsequent re-estimation of individual peptide probabilities by way of their NSP value, or number of sibling peptides, which amounts to peptide grouping information and the status of a possible multi-hit protein.
| MGMTVPDPAH | VUESQSSLPR | CAGQTVPNLAK | RTNMAGSSEGD40 | 74 | 75 |
| BAND00-20060217 | | | SSSSBD | BAND03-20060206 |
| YQGQSKMS | VPOQWNLSDQ | PARPTTLMEC | NPSQVNGSR80 | 76 | 77 |
| | | | | BAND02-20060206 |
| | | | | BAND02-20060209 |
| | | | | BAND02-20060217 |
| YQGQSKMS | VPOQWNLSDQ | BAND03-20060206 | | 80 |
| | | | | BAND03-20060209 |
| | | | | BAND03-20060217 |
| | | | | 82 |
| | | | | BAND04-20060206 |
| | | | | BAND04-20060209 |
| | | | | BAND04-20060217 |
| | | | 84 |
| | | | | | BAND04-20060206 | | 85 |
| | | | | | BAND04-20060209 | | |
| | | | | | BAND04-20060217 | | |
| | | | | | 86 |
| | | | | | BAND05-20060206 |
| | | | | 87 |
| | | | | BAND05-20060217 |

| SPOECQVAKG | GMGSPQDT | GNVRGSWAE | KMNPFPNMT120 | 88 |
| | | | | BAND01-20060206 |
| | | | | BAND02-20060206 |
| | | | | BAND02-20060209 |
| | | | | BAND02-20060217 |
| | | | | 91 |
| | | | | BAND03-20060206 |
| | | | | BAND03-20060209 |
| | | | | BAND03-20060217 |
| | | | | 93 |
| | | | | BAND04-20060206 |
| | | | | BAND04-20060209 |
| | | | | BAND04-20060217 |
| | | | | 95 |
| | | | | BAND05-20060206 |
| | | | | BAND05-20060209 |
| | | | | BAND05-20060217 |
| | | | | 97 |

| NEFQVIVP | PTLSTHDVR | QNLWAVKVEY | GLPDVILLLP140 | 98 |
| | | | | BAND01-20060206 |
| | | | | BAND02-20060206 |
| | | | | BAND03-20060206 |
| | | | | BAND04-20060206 |
| | | | | BAND05-20060206 |
| | | | | 99 |
| | | | | BAND01-20060209 |
| | | | | BAND02-20060209 |
| | | | | BAND03-20060209 |
| | | | | BAND04-20060209 |
| | | | | BAND05-20060209 |
| | | | | 100 |
| | | | | BAND01-20060217 |
| | | | | BAND02-20060217 |
| | | | | BAND03-20060217 |
| | | | | BAND04-20060217 |
| | | | | BAND05-20060217 |
| | | | | 101 |

| QHHQGKELCK | MTEDQFQRLT | PSYNADIL | SLKHLYLRETP120 |
| QHHQGKELCK | LT | PSYNADIL | SLKHLYLRETP | BAND01-20060206 |
| QHHQGKELCK | LT | PSYNADIL | SLKHLYLRETP | BAND02-20060206 |
| QHHQGKELCK | LT | PSYNADIL | SLKHLYLRETP | BAND02-20060209 |
| QHHQGKELCK | LT | PSYNADIL | SLKHLYLRETP | BAND02-20060217 |
| QHHQGKELCK | LT | PSYNADIL | SLKHLYLRETP | BAND03-20060206 |
| QHHQGKELCK | LT | PSYNADIL | SLKHLYLRETP | BAND03-20060209 |
| QHHQGKELCK | LT | PSYNADIL | SLKHLYLRETP | BAND03-20060217 |
| QHHQGKELCK | LT | PSYNADIL | SLKHLYLRETP | BAND04-20060206 |
| QHHQGKELCK | LT | PSYNADIL | SLKHLYLRETP | BAND04-20060209 |
| QHHQGKELCK | LT | PSYNADIL | SLKHLYLRETP | BAND04-20060217 |
| QHHQGKELCK | LT | PSYNADIL | SLKHLYLRETP | BAND05-20060206 |
| QHHQGKELCK | LT | PSYNADIL | SLKHLYLRETP | BAND05-20060209 |
| QHHQGKELCK | LT | PSYNADIL | SLKHLYLRETP | BAND05-20060217 |
| QHHQGKELCK | LT | PSYNADIL | SLKHLYLRETP | BAND05-20060217 |
| QHHQGKELCK | LT | PSYNADIL | SLKHLYLRETP | BAND05-20060217 |

| PHLTSDDVK | ALQNSPSLH | ARNTNVAGF | PPTTSVQYFA240 | 121 |
| PHLTSDDVK | ALQNSPSLH | ARNT | | BAND01-20060206 |
| PHLTSDDVK | ALQNSPSLH | ARNT | | BAND02-20060206 |
| PHLTSDDVK | ALQNSPSLH | ARNT | | BAND03-20060206 |
| PHLTSDDVK | ALQNSPSLH | ARNT | | BAND04-20060206 |
| PHLTSDDVK | ALQNSPSLH | ARNT | | BAND05-20060206 |
| PHLTSDDVK | ALQNSPSLH | ARNT | | BAND01-20060209 |
| PHLTSDDVK | ALQNSPSLH | ARNT | | BAND02-20060209 |
| PHLTSDDVK | ALQNSPSLH | ARNT | | BAND03-20060209 |
| PHLTSDDVK | ALQNSPSLH | ARNT | | BAND04-20060209 |
| PHLTSDDVK | ALQNSPSLH | ARNT | | BAND05-20060209 |
| PHLTSDDVK | ALQNSPSLH | ARNT | | BAND01-20060217 |
| PHLTSDDVK | ALQNSPSLH | ARNT | | BAND02-20060217 |
| PHLTSDDVK | ALQNSPSLH | ARNT | | BAND03-20060217 |
| PHLTSDDVK | ALQNSPSLH | ARNT | | BAND04-20060217 |
| PHLTSDDVK | ALQNSPSLH | ARNT | | BAND05-20060217 |

<p>| TQRITRDPDL | PYEPROSAN | TGHQHPTQ | KAAQSPSTV260 | 135 |
| TQRITRDPDL | PYEPROSAN | TGHQHPTQ | KAAQSPSTV | BAND01-20060206 |
| TQRITRDPDL | PYEPROSAN | TGHQHPTQ | KAAQSPSTV | BAND02-20060206 |
| TQRITRDPDL | PYEPROSAN | TGHQHPTQ | KAAQSPSTV | BAND03-20060206 |
| TQRITRDPDL | PYEPROSAN | TGHQHPTQ | KAAQSPSTV | BAND04-20060206 |
| TQRITRDPDL | PYEPROSAN | TGHQHPTQ | KAAQSPSTV | BAND05-20060206 |
| TQRITRDPDL | PYEPROSAN | TGHQHPTQ | KAAQSPSTV | BAND01-20060217 |
| TQRITRDPDL | PYEPROSAN | TGHQHPTQ | KAAQSPSTV | BAND02-20060217 |
| TQRITRDPDL | PYEPROSAN | TGHQHPTQ | KAAQSPSTV | BAND03-20060217 |
| TQRITRDPDL | PYEPROSAN | TGHQHPTQ | KAAQSPSTV | BAND04-20060217 |
| TQRITRDPDL | PYEPROSAN | TGHQHPTQ | KAAQSPSTV | BAND05-20060217 |</p>
<table>
<thead>
<tr>
<th>BAND05-20060217</th>
<th>NCBI N-terminal, SEQ ID NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>PYEEPRSAW TGHSPHTQPS KAAQPSSTV BAND02-20060217</td>
<td>141</td>
</tr>
<tr>
<td>SAW TGHSPHTQPS KAAQPSSTV BAND03-20060206</td>
<td>142</td>
</tr>
<tr>
<td>SAW TGHSPHTQPS KAAQPSSTV BAND03-20060209</td>
<td>143</td>
</tr>
<tr>
<td>DL PYEEPRSAW TGHSPHTQPS KAAQPSSTV BAND03-20060217</td>
<td>144</td>
</tr>
<tr>
<td>DL PYEEPRSAW TGHSPHTQPS KAAQPSSTV BAND04-20060206</td>
<td>145</td>
</tr>
<tr>
<td>DL PYEEPRSAW TGHSPHTQPS KAAQPSSTV BAND04-20060209</td>
<td>146</td>
</tr>
<tr>
<td>DL PYEEPRSAW TGHSPHTQPS KAAQPSSTV BAND04-20060217</td>
<td>147</td>
</tr>
<tr>
<td>DL PYEEPRSAW TGHSPHTQPS KAAQPSSTV BAND05-20060206</td>
<td>148</td>
</tr>
<tr>
<td>DL PYEEPRSAW TGHSPHTQPS KAAQPSSTV BAND05-20060209</td>
<td>149</td>
</tr>
<tr>
<td>DL PYEEPRSAW TGHSPHTQPS KAAQPSSTV BAND05-20060217</td>
<td>150</td>
</tr>
<tr>
<td>SAW TGHSPHTQPS KAAQPSSTV BAND05-20060209</td>
<td>151</td>
</tr>
<tr>
<td>SAW TGHSPHTQPS KAAQPSSTV BAND06-20060209</td>
<td>152</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SRANPQGQQ TQMQVLQLL</td>
<td>153</td>
</tr>
<tr>
<td>PX</td>
<td>BAND01-20060206</td>
</tr>
<tr>
<td>TEQRLQ DYPQVLQGTS SR</td>
<td>BAND01-20060217</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND02-20060206</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND02-20060209</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND02-20060217</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND03-20060206</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND03-20060209</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND03-20060217</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND04-20060206</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND04-20060209</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND04-20060217</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND05-20060206</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND05-20060209</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND05-20060217</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND04-20060206</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND04-20060209</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND04-20060217</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND05-20060206</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND05-20060209</td>
</tr>
<tr>
<td>PETEDQRLQ DYPQVLQGTS SR</td>
<td>BAND05-20060217</td>
</tr>
<tr>
<td>MGESNSCSCI TLOTNGEKEF MTDPSVARR WGERESKFKFM</td>
<td>160</td>
</tr>
<tr>
<td>MTDPSVAR</td>
<td>BAND01-20060206</td>
</tr>
<tr>
<td>MTDPSVAR</td>
<td>BAND02-20060206</td>
</tr>
<tr>
<td>MTDPSVAR</td>
<td>BAND03-20060206</td>
</tr>
<tr>
<td>MTDPSVAR</td>
<td>BAND03-20060209</td>
</tr>
<tr>
<td>MTDPSVAR</td>
<td>BAND04-20060206</td>
</tr>
<tr>
<td>MTDPSVAR</td>
<td>BAND04-20060209</td>
</tr>
<tr>
<td>MTDPSVAR</td>
<td>BAND04-20060217</td>
</tr>
<tr>
<td>TDPDEVARR KEFEMB BAND04-20060206</td>
<td></td>
</tr>
<tr>
<td>MTDPSVAR</td>
<td>BAND05-20060206</td>
</tr>
<tr>
<td>MTDPSVAR</td>
<td>BAND05-20060209</td>
</tr>
<tr>
<td>MTDPSVAR</td>
<td>BAND05-20060217</td>
</tr>
<tr>
<td>NTOKLSRALK YYTDENNMTK VQHPYKAYKF DPQGIAAQGQ</td>
<td>400</td>
</tr>
<tr>
<td>F DPQGIAGAQGQ BAND02-20060206</td>
<td>178</td>
</tr>
<tr>
<td>F DPQGIAGAQGQ BAND02-20060209</td>
<td>179</td>
</tr>
<tr>
<td>F DPQGIAGAQGQ BAND03-20060206</td>
<td>180</td>
</tr>
<tr>
<td>F DPQGIAGAQGQ BAND03-20060209</td>
<td>181</td>
</tr>
<tr>
<td>F DPQGIAGAQGQ BAND04-20060206</td>
<td>182</td>
</tr>
<tr>
<td>YYYDENNMTK YAYKF DPQGIAGAQGQ</td>
<td>BAND04-20060209</td>
</tr>
<tr>
<td>YYYDENNMTK YAYKF DPQGIAGAQGQ</td>
<td>BAND04-20060209</td>
</tr>
<tr>
<td>NTOKLGR</td>
<td>BAND04-20060217</td>
</tr>
<tr>
<td>NTOKLGR</td>
<td>BAND05-20060217</td>
</tr>
<tr>
<td>HPFPRESSLYK YPSDLPYMGY YPHQRPMNF VAPHPALPRT</td>
<td>440</td>
</tr>
<tr>
<td>HPFPRESSLYK</td>
<td>BAND02-20060206</td>
</tr>
<tr>
<td>HPFPRESSLYK</td>
<td>BAND02-20060209</td>
</tr>
<tr>
<td>HPFPRESSLYK YPSDLPYMGY YPHQRQ</td>
<td>BAND03-20060206</td>
</tr>
<tr>
<td>HPFPRESSLYK YPSDLPYMGY YPHQRQ</td>
<td>BAND03-20060209</td>
</tr>
<tr>
<td>HPFPRESSLYK YPSDLPYMGY YPHQRQ</td>
<td>BAND04-20060206</td>
</tr>
<tr>
<td>HPFPRESSLYK YPSDLPYMGY YPHQRQ</td>
<td>BAND04-20060209</td>
</tr>
<tr>
<td>TSSSSPAADN PYWNSPTGDI YPMTRLPTSH MSHSLQTVY</td>
<td>479</td>
</tr>
<tr>
<td>NSPTG</td>
<td>BAND02-20060217</td>
</tr>
<tr>
<td>SPTGDI YPMTR</td>
<td>BAND04-20060209</td>
</tr>
<tr>
<td>SPTGDI YPMTR</td>
<td>BAND04-20060209</td>
</tr>
<tr>
<td>NOTE:</td>
<td></td>
</tr>
<tr>
<td>E-BAND×• represent ERG2 peptides in ERG1 experiments</td>
<td></td>
</tr>
</tbody>
</table>

The table shows the coverage map for ERG2 obtained over 3 different experiments. The underlined amino acid sequence corresponds to the in silico translated sequence of ERG1 that was cloned from VCAP cells. The amino acid sequence GGAEIFPNTSVYPEATIQRTTRP (SEQ ID NO:196) corresponds to the exon that is specific to ERG1 and is missing in ERG2. The remaining amino acid sequence correspond to ERG2 sequence identified in each of the three experiments. ERG2 was identified in Bands 1-5 in all the experiments. The peptide sequences for ERG2 obtained in each of these bands is illustrated. A very high coverage of the ERG2 protein was observed over the three experiments. The coverage map showed that the coverage of peptides in the N-terminal region of the cloned protein, corresponding to the first 50 amino acid residues were rarely observed in the mass spectrometry coverage map. However, the peptide VPQQDQLSQQ (SEQ ID
NO: 197) that starts with aminoacid valine was found to be highly abundant and thus identified in all the experiments. Closer evaluation suggested that aminoacid in the 47th position was an in frame Methionine. The lack of any peptide upstream (Nemrivirus) of the 47th methionine in multiple experiments confirms that it is the N-terminal aminoacid of ERG2. Further, the presence of an Arginine residue at the 50th position makes it a potential tryptic cleavage site. Digestion by trypsin at this site would result in a shorter N-terminal peptide MSPR, which is too small for identification by ion trap mass spectrometer and a longer C-terminal peptide VPQDWSQP (SEQ ID NO: 198), which was identified in all the experiments. Also the peptide sequence MIQTVDPAA HI (SEQ ID NO: 199) was identified in a single experiment at a very low probability score. This maps to the N-terminus of ERG as reported in NCBI. This sequence was not a part of the ectopically overexpressed construct that was cloned from the VCAP cells. This could have been obtained from the in vivo ERG that is expressed in PHINX cells and thus may represent part of the ERG associated with benign cells. Thus, in summary, the results indicate that the third Methionine is the translational Start site for the TMPRSS2-ERG fusion product. MASTIKALE S VVSEDSLIFE CAYTPHPLA TEMTA YGQTSMPSR VPQDWSQP (SEQ ID NO: 200)

The First Methionine is the translational START Site for endogenous ERG.

FIG. 20 shows a schematic of the endogenous and fusion polypeptides.

EXAMPLE 18

**FISH Analysis on Urine Samples**

To isolate and prepare prostatic cells from urine, ~30 ml of urine is collected following an attenteive digital rectal exam. Immediately, 15 ml of PreservCyt is added, and the sample is centrifuged at 4000 rpm in a 50 ml tube for 10 min at room temperature. The supernatant is discarded, the pellet is resuspended in 15 ml of 0.75 M KCl for 15 min at room temperature, and centrifuged at 4000 rpm in a 50 ml tube for 10 min at room temperature. The supernatant is discarded, and the pellet is resuspended in 10 ml of a 3:1 ratio of methanol: glacial acetic acid. This is then centrifuged at 4000 rpm for 8 min. The supernatant is discarded, except for 200 μl, and the pellet is resuspended. The resuspended pellet is then dropped onto glass slides and allowed to air dry. Hybridization and probe preparation are as in Example 2 above, with the ERG 5’3’ and TMPRSS 5’3’ probe pairs.

All publications, patents, patent applications and accession numbers mentioned in the above specification are herein incorporated by reference in their entirety. Although the invention has been described in connection with specific embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications and variations of the described compositions and methods of the invention will be apparent to those of ordinary skill in the art and are intended to be within the scope of the following claims.
tctggtacca actgctcatc attgtc  26
<210> SEQ ID NO 5
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 5
tcaggtacc tgacaaatgt gacgcag  26
<210> SEQ ID NO 6
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 6
catggagtgt gggtctttt cttg  24
<210> SEQ ID NO 7
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 7
aacagccctt taaattcagc tatgga  26
<210> SEQ ID NO 8
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 8
ggagggcctc atccccactt g  21
<210> SEQ ID NO 9
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 9
ttacocccatg gacccacagat tt  22
<210> SEQ ID NO 10
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 10
cttasagct tgtcgtggga ag  22
<210> SEQ ID NO 11
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 11
cgcaggtta ttgctgcagc agat  24
<210> SEQ ID NO 12
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 12
ccatattcct tcaacgcca cttcc 24

<210> SEQ ID NO 13
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 13
cgacttgagc aagggcacac tga 23

<210> SEQ ID NO 14
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 14
catggactgt gggttctttt cttg 24

<210> SEQ ID NO 15
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 15
gggttccgt aagccactc aa 22

<210> SEQ ID NO 16
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 16
cctggtcgag ggttgagaca 20

<210> SEQ ID NO 17
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 17
taggccggs gtaagcagg g 21

<210> SEQ ID NO 18
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 18
gtggcacac tcacacaacg actgg 25

<210> SEQ ID NO 19
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 19
cgcggctaa gcagggagc 19

<210> SEQ ID NO 20
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 20
ccggcccatga aagcacaac tt  

<210> SEQ ID NO 21
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 21
ccgagtggag cggagtga

<210> SEQ ID NO 22
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 22
cgggcgattt gctgctgaag

<210> SEQ ID NO 23
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 23
gccgccccctc gactctgaa

<210> SEQ ID NO 24
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 24
gagccagctc tcctggaagt gact

<210> SEQ ID NO 25
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 25
cctggccggtt otcttggagtc

<210> SEQ ID NO 26
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 26
cgggccgggg astggagt

<210> SEQ ID NO 27
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 27
cctggaggt accggtttgtc a

<210> SEQ ID NO 28
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<210> SEQ ID NO 29
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 29
aaataagtt gtaagagggg cctcagcact

<210> SEQ ID NO 30
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 30
atgtaaaga gtttttttoc ccgc

<210> SEQ ID NO 31
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 31
gaaagggctg taggggcccac tgt

<210> SEQ ID NO 32
<211> LENGTH: 327
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 32
cctggcgcgc cgagggaggt agtgcgaccc ggtcgggctg gcacggcctc ggcaacgcgcg
ctgcaacag ogggcgcggg caacgccgaa cacacacgtg cgggacacgc cctcccccga
ctggcggctg aacccagcgg ctattccagg atcctttggcg aacccggaga agccgcttgtg
acccaaacga agacaatggg ctccagagaa aaaaaatgag aggaaaccag gcacactaaa
gccgcgcagtt ttcggagagc tgggtatagg gcgtggttac tgaaggacat gattcagact
gtcgcgcacc caagcgcgca tataaag

<210> SEQ ID NO 33
<211> LENGTH: 6158
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 33
gttgtagaa gcggctagcc ggaggaaact tcctagctaa tctgctaaat ataaaatcgc
agagctagat ttcggagagc caagcgcgtg gatggatatt agaaccagca agtggttacc
tggtcaca aatgtcagcg tggggagaaat tgtaacccga aacccaaacaa tcttcggaaa
agaaaaatc tcctagagaa tctgggctcat gattcagag aacttccttca agatctaaat
caattacag aacagtcgtc tgctagagct caggtacttg agaatagtaa gcagtttgtgta
ccagaccttc atgcgtcggat ctgggctctt ctggtgccga ccctgaatct caagaaagac
ccccagctc catggtggag aatcggcatg gcgtggtta cagagacaccc cttaaaatcc
agctagatg aaaaagcgcct gctcaaggcgtag agtggttactg acagatgctg caagtttgga
agagctagat ccaccccccac caccccatgc aacagcgcag tctcccccact gcattgtgca
-continued

tatgctggttt ttgtgatgtc atatctgttg cgttttttt gttggttttt cgtttttttt tcgtttttttt ctttttttttt tcttttttttt cttttttttttt ttttttttttttt tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt...
cont-continued

```
ggaagaatc attatggcag aagctacagc attttgccag atgatagggga tttctaatat  5400
gagccacttt gctctacag cagccctttc agagcttgta atgagaaac attacagag  5460
agaggtctt ttaggtgccgt ttagcttggag aatccagaaaa cccaaatact aaatattaaa  5520
atataaagc agaatgatagt tttcaccatt tggctttgta tggagagagc aggaaattaga  5580
attataaa attaatcaac tttggtaaag tggatggtgga aaaaatattaa agaagggaaa  5640
tggtacatatt aacagaccaact acaagagatt ggaaatataca acagttccaa aagatgttaa  5700
agttggtccct gatactgaa aatcccccac agttttttaa gatcctcaac actagttttta  5760
ggcttttcttt ttttatatac atgagaaata cttgacata cttttttttttttt aaacctatca  5820
tggaagaatcc ggttttaagcg gagaacttgga ttagaccaag ttagataatttt cttagaataa  5880
ttagattgct gagaatttttc gattacagct ggtagcatgttt taggtgattc  5940
atgatcaagc tagggccttt tttcctatga gctctccgta ttagttctct ttttttttaaa  6000
attatatatttg tgggagttta ttttaatact acagtgggttgt attaataatta  6060
tacctgcgct tggagagcaca atggtatagaa aatcattgatt atacagaaat aatgtcgtttc  6120
aaaaatatatt gataatacaga aagcggcaag aactgaaaaa  6158
```

<210> SEQ ID NO 34
<211> LENGTH: 5228
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 34

cgggccaggg cgggacagca ggaagcgggcc cggcggaggg cgggttggga cccgagcggg 60
cctccctcgc cggaaagcgc caccctcccc ccggaggcgc cgcgggggcc gggcggcagc 120
gggggaaaaa aagccgcgag gaattggga tttttgggaa gaaagcgga tttcccccogt 180
ccccctcccc ctcgtttatc cttctcttaa aaaaacaacc aacagtaatc gcacaccttc 240
taccatccgc taccgctccc actttgctgc ccatcagagc gggcgtgctgt cttcaacgcg 300
tccattacac tattaagctt cttttccccc cccgtgatttg 360
aatgtgagtc gtaacactca tttactccaa gcgaatagaa aatgatgtctt caagagattt 420
aagctacttt cagttggcttc actaagaaac aagcaagcact ggggatccca aagagccccc 480
ggcaagtggac aagcaggccat gttcggcagct tggattgcgtt ggtgttcaaat gatacagcc 540
tgaagtggt gtcttttctca aagcttgagc aagccttcttc gcctctgtga 600
aagaactgtc cttttggtgc gctcagacgt tttttttggga ctccttttaa gccttctaatag 660
agatctggca gaaagagactt gttgaaccaat aacagatata tggagtctaa cccagctact 720
cgaatccgct ttttatccct cttctttctg tttggtggtt gcctggttgg 780
ttcaccacat gcggcttccag gacgccgaag tctcactcag tggatccactg 940
cctcagctc caagagatcct ttttcttcac agatagcgaa tctactccccc ttgctatcctt 900
tggagagcc ttcacgacagc agacctttgc agatagcata tttgtatata aacaaagaagc 960
tggctcaccgc agcaacacat tggagagggc gcggacgtcg tggtaaatcct ggggggcagg 1020
acactcttta aagcaagagat cgtagatgctt gttgcttgctc ttacacccag tocctgaga 1080
gcccagcttc ttatcaccag ctggagcttgg tttttctcacta tggagacgctt gtcgcttaca 1140
acatcggcct gcctctgctcc aacagacag ccaagggcag cttccagagac ttaggtcgggg 1200
acagtctgca cccctctaag gagctagcctg ttctttctctc tggatggtgga 1260
cggcagctg gacccatcag cttggcagtt tttctttgga attactcact gataaatcc 1320
gtcagtcttt tatacgcttg gacggagatg gctgggaatt caaacttctt gagccagatg
aggtggcccg gcagtgagga aagagaaaaa caaaacctaa gatgaattat gagaacctga
gecggtgccct agcgtacctt tacgaaacaa caatcacatca caacagacagc gggaanaagct
agctgacccct ttggtgctgt gcagtgcaaga gcctgtcggg gcctagcttg gggtaagccg
dcgctagtgt gcacgtccgg cagctgatcg gcacgtgaagg ggcttggggaga
acgctgctga gaccccctcga gacagcctgt gttgggtgga cttggatatt tggagttta
cttcattttat tattttttaagg ccattctttcag gggtaaggaga taagacgtggt
ccaggttga ctaaattgct gcacgggtaaa gctagccgaa gctagccgaa
ttcagggta gacatctatt tcttccgaga cattcgggaa cagctaggtg ctggtgtgta
ccccatttta gcttacgtccg cattactgtct ccattctttcc caaatttattttttt
tctggctccta ctaaattgct gcacgggtaaa gctagccgaa gctagccgaa
ttttttaatttttttta tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
ggagcgctc agggagtgtt ggttggcaact gttttgacc tattggattc ttagattta 3700
tgtgatatcttgttcgcgtc gctgtagtga atgtgtgcactgtgcgtgaat 3840
gttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtg
-continued

tggcagagtc tgtgttaacc tgggcaagga aegcttttcg gagctgggca acgaaccttgt 660
gggtgacatt cttcgggcag atctggcagca aagtacaaag gaaaaaccaag aaagagcaga 720
agatcaatg gaaagaaatt cacacctcctactcttgctcg cattggtatg acacgaatac 780
attagtattt ggcgcagcag aggcgcctta tggatcagc acacagaaatt acaccccaag 840
cgcctcctag gacagcactgt gtccggcctc cccaccccagc gatacactgt gtacggcaga 900
gtctcaagtc ttcocctcagc ttcgctcagc gtaactacact ctgcctgctcag 960
tccaggtcct cccaggcagca acttgaatgt gctcaacaca aatcttgagga ctcccacaga 1020
ccacagcaccg ccagacccagc tggagctgagtg caaaccagttt atacctgcag ctgctgctgc 1080
cgcctgacta gcaagggcgct ctattacagt gttgcaagttt ttcctggcagc gtgatgtcgaag 1140
caacacgactcg ccacagcactgt ctacgctcagc ccagacccagc gacgctgctgctc 1200
cccgcggaag ccagacccagc tggagctgagtg caaaccagttt atacctgcag ctgctgctgc 1260
cgcctgacta gcaagggcgct ctattacagt gttgcaagttt ttcctggcagc gtgatgtcgaag 1320
caacacgactcg ccacagcactgt ctacgctcagc ccagacccagc gacgctgctgctc 1380
cgcctgacta gcaagggcgct ctattacagt gttgcaagttt ttcctggcagc gtgatgtcgaag 1440
caacacgactcg ccacagcactgt ctacgctcagc ccagacccagc gacgctgctgctc 1500
cgcctgacta gcaagggcgct ctattacagt gttgcaagttt ttcctggcagc gtgatgtcgaag 1560
caacacgactcg ccacagcactgt ctacgctcagc ccagacccagc gacgctgctgctc 1620
ctctgacctg cgcctgacctg ctctgacctg cgcctgacctg cgcctgacctg cgcctgacctg 1680
agacggcagaa aagggcagca tggaaaatgt ccagacagtt gcagccagcag cagctgtgctc 1740
attgcaatcc caacgcagcag ctctgacctg ctcgctctcc ctgtgacgcag aagggcagca 1800
ccaattctca tcaaacgtgct tttgctgaaa aagtgcagag aagggcagca tggaaaatgt 1860
tctctgacctg cgcctgacctg cgcctgacctg cgcctgacctg cgcctgacctg cgcctgacctg 1920
ggagagcttg gttgaacacc accagactat tttgattttt tttttttttct tttttttttct 1980
ggagagcttg gttgaacacc accagactat tttgattttt tttttttttct tttttttttct 2040
caattgtaga gtttttttgtt atatatattat atatatattat atatatattat atatatattat 2100
acaacgtttgta ttttacaaaa atttgtatatt gtaatttttt atatatattat atatatattat 2160
gaataaagtt ttcagacactg tggagggtttt tttttttttttt tttttttttttt tttttttttttt 2220
tgcaagattg taacgtgcagag agggctgactg caaagatgtg tttttttttttt 2280
attaatgtg gttttaaaatttt tttttttttttttt tttttttttttttt 2340
tttttttttttt tttttttttttt tttttttttttt tttttttttttt 2400
cttacatgtt ggggggggtttt tttttttttttttt 2460
tctgctcctgt tttttttttttttt 2520
gcagactgct ctacacttctgc tcaacagttg tttttttttttttt 2580
tttttttttttt tttttttttttt tttttttttttt 2640
cagactgct ctacacttctgc tcaacagttg tttttttttttttt 2700
caacacttctgc tcaacagttg tttttttttttttt 2760
ttacctgtaga gggaggtttttt gttttttttttttt 2820
agtttacacag agtttacacag gggaggtttttt gttttttttttttt 2880
tttttttttttt tttttttttttt tttttttttttt 2940
cagactgct ctacacttctgc tcaacagttg tttttttttttttt 3000
cagagacctt ttctgaacgc agagccggaga cttaaagttgg gasagagaaaa gcatcggagc 3060
catcatctc gagaagaagtt tttcatcaca attgagacctt ttgctaagct ttccaasaagag 3120
cagctgacct acgtatcatc cccgcttctaa tattagccccc gctcaagttg gtttcaaatgt 3180
cccagcgccttt gctactgctg ggccctggtgag ggttgtgtgta cagcttgtaaa 3240
gacttcctct gtgaagacac tgaggatactg ggtggtccgg cagctgctgga agagcaaagcc 3300
cagagctgcgct ctgctgctat aaccacaaaa gacacttccc agtatacata aacagacagt 3360
gtttttctca aagaggatgt attatcacct tgtagactcct tttaataat atacatgacat 3420
gttactggaga acatccgctg gcacaaagaa ttcctagcag tgcgtccttg cttgtaagtt 3480
tgaaacgccgct gaaatgacca atataggctc gggctttttac tacaagagac ggaattctag 3540
ggggtcagcc taacagctgta gaaacactca gtccttttaa aaccattggca tggtaaataa 3600
cagatctcata gataaaatct tttgtaattg ggcctctatc ctctcaattaa taaggtattt 3660
tgttttatata aa 3672

<210> SEQ ID NO 36
<211> LENGTH: 2668
<212> ORGANISM: Homo sapiens
<400> SEQUENCE: 36

cagctacacac aggtaccctt ggagatggctg gacgaacacca cccagcgatgg acacatcctgt 60
gagctgttgg caggttctcg tgcagcgtctg gagaagacaa ggcatastgc acatcctctc 120
cttgactcca cgggatcttg tgaatctcaca ggtgagtggat ggcagagggg gttgcgctgg 180
gtgattgctca gcaacaagca aagcataaat gcctttgacag ccgagccttc ggccttcg 240
gactactat gacaagaacca tccatcgcacag ggtgagcccg cagaagttgct tttaataatt 300
ttgttgtctct cctgagctcg cccgctgctc ccctgtgagcc ctccttgcagcc 360
gtttgtttct ctccacatgc caaatgtggc cctgtctgct atacatgcgc cccacgaggca 420
cactgtgttc cgaagacgac gcacacccaa gggggcgagtt gtcagacgcc cggcggtgctt 480
ggcagacaac agcggcgcag acatcctgcag ccctgctgcc tggctctcctc ccacctgcagc 540
tctggcaggg gcagccagcgc cccctctgctg tggctctcct cttggtctgc ccagttgaac 600
ggcctgtcc ggagcttaag aggcttgtgg tggctctgct ggtcctcagt ctattcctga cccgctggca 660
ggccccacaa ctgtaatgagc acacatgcttct gttttgaggg gcgtcggctgg 720
gggcctgctg cagctgctgctt gaccgggttc ggtcctcgag ggtcctgctg 780
cccagaagtg aagtagagac gcacccagga agagttggaa gttctggaagg aggagaggtt 840
tgtggcctgg cccccgccgc cccagccgca attccctca ctgagccggc gcacggaggg 900
gctggtcggc cttggtatgg acacccggcg ggcgtccgct gcacccagcc cttggagctg 960
tgcagatcct cccgaggcag cggcagggga ggcctgtgca ctgagctctcc cttggagctg 1020
gagctgctga gggctgctgg gaccgccagc gaccctggca cttggaagtct gcacccagcg 1080
ccagcccgcc gcgcggcgcgc tgaacccactc cctgctttctc acgcacactc tcagagccct 1140
gctgctgca ccacagtcgct ctctcatcct ccacatcctc ctgagctctgc gcaccctcgt 1200	tgccggcgtg agcggcgcaca agtcctccct cccgttctcc ctgagctctgc gcaccctcgt 1260
gccctctct ccctcactgc cccgctcgc cccactccct ctgagctctgc gcaccctcgt 1320
cccaagcaca tcgcaaacac accggccttc cctgggcttc ctgacctcgt 1380
tctctctcg cccgcctact ccagagaaaa ctcagtttca ctgaaagact ctgctctga 1440
tttgtgggag gtggggagcc ttggagaaga ttactccccaa agaacaactct cattatcttc 1500
tcacaagaa acaccaagct tccacaactt cctgttttttc tgtcagcccc ccagtggccg 1560
ccttttaacct ccacagctta cattgctgag ggagggttatt ttatttatatt tttggaagggc 1620
acttgaggag gctcactaca accttttagg gtgggatagg catcctcccc ccaccccccac 1680
tttttctccc aacagcagat aacagggctg tggcctgaga acgaccccc tttttttttt 1740
tcacaagctg cccctgggga gatgaggagg cccctgctgg ctgttttctgg ctgtaggaaga 1800
gagatattaag gttgttggttt attatccttg ggctacctca ggggtgccag gagaattttgt 1860
accattttaat ggtgtggagag ttggggccac ggaagatcaca ccctctggag atagaaatttt 1920
cccctccccca ccaccccttc tcaaggacgtct tctctttttttt aaccacactctt ttggcagcggg 1980
aggaattgtcc ctttcttttt ttccccctgtg aagccatctcc ttttgtcgcg aacctctcctg 2040
gggctctgccc tgttttttccc caaggtaggg tttaaaaaaggg ggtggctccc ccgctggtggg 2100
gccctctatg cacaggctac gggctctccc tcttttcttg gctgtctcttt 2160
aatccatatt taaaagggcc ttccccatag ggggagggag gctactctct ccatataattg 2220
gtgagtttggg tggggagaaag ggggtttttg ggggaacttt ctctgcgcct ccoccaaccc 2280
agacatttttat tttgtgtatt tctcagataa accaccttac tttcagcttc cctccctcaca ggggcccacaat 2340
tctctgaggg cggagacaaa ggacccctct aacgtctccc gatgctgagg ggaggaatgg 2400
ggacataaaa gcctctccct gcttttattg aacggacaaga gcaggggttt aaaaagactct 2460
tggggctacag tgtttagctt ggcccaagcc accgtctgggg acctgggggt tggctatttg 2520
ggggaggtgg ctacagcttg ctccccagtt ctgttttattc ccaccaaaatg gaccccttttt 2580
ccccctaaag cgttcccagag aaggggagat tttgctgtgta aatatatttt tttccaaatg 2640
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2680

<210> SEQ ID NO 37
<211> LENGTH: 5992
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 37

ggcgcgggg cgccgccgcc gcgcgcgcga ctcgcaaccc cgccgcgccc gcccgcgcgc 60
ttcgcgagc cggccgcggc cgcgcggcgg ccgcccgcgc ccggcgcccc 120
ggctggcaga cccggcgccc gcgcgcggcg agagatctct ggaaaccaat ctttactgc 180
cggcgctgg ggcccggcttt gctcctcggg gttggggaga gggagaagaaa gttggagaaa 240
cctgagactct tctctctctg tctgtctgtg agacactctt gactcactct ctcaggttag 300
cattaagcg gacaaattt cattacaccc tccagcagtc ccaggccgca gttacgcttt 360
tcgagacgc ttcctagttc caagtctcct aggcgtcaggg aatggaggaag actgcattcg 420
tctggtcag cactctctct tgtcagcacttt ttagactggg aaggtgtcgag tgccttgtg 480
gctcaagtt gcctaaatttt gttttctttt gtaagccaaata cccggcgtgg cggcgcgcggc 540
gactggcgcac gtttcctccct tgcgacacag cattatgctcg ggcatatgctgg ggactgttg 600
aggtgatgtg ctcttgggtaa ctggctggcag ttcgggttggg agagatgctt atggctttttc 660
ccctccatct tccctccacct ccgacaggcc cttggagaga atgtattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
aggtgaagt acctagacc ctcgaccttc ctctagttga gtcgacagctg 3360
taggttgctt gcctcagcgg caagagaaat ttctccctgg cctctctgcag aagggacca 3420
acctaggctt cacacaagat gcacagagCc gacagacgct gacagacgct gacagacgct 3480
aggttccctg accttgctctg gcttgctgct cttctctcag gcacagagctg acacagggc 3540
ataygcggcc acctccctg gtcgccccgc cccccaccat ccccccctgc ggcaggtgac 3600
gaccttcat atagggcagc ggtggtctag ctcgctttgaa cttgggtatg aatgcttgg 3660
atctacgccc gctgcaggttg ctctctcttg gcacatccgt gacatgctgc atctctattt 3720
caagctttca gatactgttg atctctctct atagtctcagc ctaactacag attacagtgc 3780
attctgaa gtcacagtatt ttaagaaggg agagaaggg atttttactc gcacccctct 3840
gtatgaatat gaaatcagag accaggcat gctggtgtca ggcattagc atctctgctg 3900
aggctcttca ccacagtgcc aagaacctgct cccacactgt ctacaaactt ctaacaaggtc 3960
aattgtagg gggaaactcc ttgctgccca agaatcttta tatataacag aatacatcag 4020
gaaatagtt ttagctacagg atgtatcctt cagcccacttg aagaaggaat aagggcgcct 4080
cactcaata tagctctat cttcaccattg ctcgctttcc aatctgaaca 4140
gttttaccc cccaaatggg cccaatccaa cccctctctg agctctagac acacagtgctg 4200
gaaatctct cccactctcc cctctctctg cctctctctc aacacacgct 4260
ttatatatga aatatcctca aatcatctaa aatatttacct ggtcctctgg tataactgt 4320
atatagccaa taaatagatt tgtgatttta actttggtgc ctcctcccaag aatctctat 4380
ttatatatat atatatatat atatatatat atatatatac aacaacacac 4440
acataccaca atattccagc atacaacag calccatattc caaactccag cccgacaccc 4500
tctgtctcttc aacaacctg ctaaatctgta aatacatctatg cagaccttctg 4560
aattccctta actgcctagtc ctcctctctt ttgctcttta ttagacacag gcggctagtc 4620
cacgcctcttc atacacgtag ccacaccacta gttcctatag tcaggtagca 4680
aaccttcttg gaggactcgc tcggactctag ggcacaccag gccattacct gctgtatag 4740
tgcttctcag gtttatataa tagctccctca ccaacaccc cgtcctcctg ctccacacta 4800
agtaagtggc gtgtgctagc gagacacactg taagctcag cagaggaact gcacagggac 4860
cocgcaccttg gatgccgcttg cctctctgctt cagctctctt ctacatctct 4920
aagctctttcc gcggcctctcc tccatctctg cagacacagc cggatagcttg ctgactctgg 4980
tatggttag ctaatctgaac cttgcatagt tggacctacg gacagctttgt tttgtttaag 5040
tgcggttca tctgcgaatt gcactgcttc tgcgggagtg ggtataggctt ggtgtttgag 5100
auaatagtt cccctctcag gctggttgtt cccagacctc ccacactacag tcctctgttg 5160
gtcctttctg actgtatctt tgtctcaattgt agaatcgctg tgtctctgttg atatcatttg 5220
ctgtctctctg tcctccatgat gataggttaa ccacttctcc ctcctctctc tgcctctgtg 5280
ccacagtgcc atataatacct gaaatagggg ccatactttg cccacaaagc aacttccagt 5340
cagggaggtg aagttgacct gcctcgaggg ctccacagcag cacaagagtt gattccgctc 5400
aggttcctca cagcttcttc ctcctcttcag gcgcagcag ctcctgttgctg 5460
accgggtgaa cccactacca gggacagggc acgttctcttc agatttcagct 5520
agttggagtc accttccttc ctcgctttctt ctctctcttt cccacagcag ctgttctctg 5580
agggagctc ccccgctgtt gggacaacct atacacccct cctaaaccc cccagacgcct 5640
aacctctgatc ctctctctgg cccctctgac ccacactacag aagttgacct 5700
tttcacacct gcgaagtgaa cattcctggt gattttctct tgtggcgcgca accaaactgca 5760
aacgagacg acctcagcat ttcttcccgc ggcccccttt tgttttatttt ttaaatcctct 5820
ttgtatatgcct tgtggtggct ttaaggggca cgaatcaaat gaaatcaca aaaaaaagaa 5880
aaaaaaacaca aaaaataacct ttctctcagg gcaagctgag tgtcgaatatt tttaaagaa 5940
aggtctaatct tacacctct tacatatttt ttatatatgaaaaa aaaaaaaaa aa 5992

<210> SEQ ID NO 38
<211> LENGTH: 1879
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 38

gattctctcc cacgagacct tcctagttct gcagcggaggt taggggttgg gggagggag 60
actcgggggc ggggctctcg gggccacgca gcacggcggca gggagagctg caagcagaa 120
agggctgggc ttctctacct ccacatgtgaa cgaagggaggt aagatcagct attcctccta 180
taaatcttgc ggacgctcatt cctcctctag gccacacagt gcacagccaga tgtgagactc 240
aaatctacg ccgggtgtaaa ggggagactc gcagtgcgca gtagaagctcc aggctcagc 300
cgccagttag gaggctgagag ccggctgcgc actggtgtcgg ctggggagag cagggagtct 360
cctgctcacg ccacgcaggg ccagcgtgtct cagagtaagc agcggcctc agctcctcca 420
ccagggca ccgcttggac ccgtgcagccca gttcaggtaga gctctctctgt gcaggtgtcct 480
agttacacaa gccggctagg ccagcgctgg tgttgagacc cttttttgga gggatcttca 540
gggtgagacg ggcctcagcag cactctctccag tccccccggga agaggtgacct gggccccttc 600
agatggaccc cggagggcg cactctgtgc aggccacagc cccaggggctt acacagcaact 660
tggccacac ggtgaacocct gcggctggcaac gcggcagggg ggtgctgtcc gcctctctca 720
actattgcta cttgctgagc cttgctgcgc gggccccaggg ggtgctgtcc cttccccggca 780
tgcctccgcc ccctcttgag gcgcagtctag ctgctctgcg ccggctgtgg gattaagctgt 840
atcagcgtgct cctgtgatacc cgtatagcgc ctcacatctaa cggggaggac aagagccgca 900
agattctctcg atgttggttag ccagatgggcc tggcagcattgc tggggggatt cccaaagacc 960
ggggtgaact gcgactcaggg cagatgtctct gcggctgtcgc cgcctttgatt aagotattaa 1020
tcattaaaga ggacccgggg ccaaattctcc tgggtcgatt ttcgaaagact cggggggaga 1080
tggtgctagg caacagcagc caagtgagcc gcggagtagg cagagaaatc 1140
agtttaaggg ccagcagca gaattctctcc cgtgagggag ggtgggagac ccgggaccgg 1200
gtacagctgg gcagcgctgac gattctccca tgaagggagc ctctctctcc cccgagcgca 1260
gcggagccgc cagccagact ccctgaccacag aggattcagc cctatgttgg gaaagctgg 1320
gggcttccg atccaggaca ctgggggcag caggtgcata ccctggcgc agcatctctg 1380
gggtactattg cctgcagcag gattctgagt tggcgctctgg ccggctgtgg aagagagcc 1440
agggcacgt ggagagcgca gcggccaccc gcggcagctg cttgaaagtg attcaccagg taggggaat 1500
ggaagttcttc atctgagctg ctttctaccgc eaacoccttg gggaagagcc 1560
gaaaaagttc taggtctttcc cgggagctag gcagctgcct gtaagcaattt aaagatgtgt 1620
agctggaaaaa aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1670
<400> SEQUENCE: 39

gagtcagcc gcctgtgccg ggagcggttc accgtcttcg gagcggcttc gcaccagcctt 60
tgccacacc gccaccgcgc gcctgcgcgc tgtctgagcg cgcctgcgcgc gcacgggcgc 120
cctcaaggg agggagccgg ccggcctcag gggatccttc ctcccccaca aattactcaca 180
tgcgcacagg ttctcataag gcgttagaag acatcaggac agtactttga 240
tcagcaagtc ccttctagtt cctccgagga atttctagct gaggaattga gaggcgggccc 300
tgctgtgagcc agaaagagga aagtttccag cacagactcg gctcagcaat ctagagagct 360
attcaggt gctcagtcac ttccagaggc tcgtgtagct gaacacaaag tctcagagta 420
tgcagagtt gctccagatt tcagctgata taacctgggtg ctctcagcc caaccccaac 480
cacagtacaa epggagtcgc acacccctcc ctctcagctg tgtccttgta gcacagcrafted 540
ggtcttcttg gctaatcagt gagaagagtc ctctcagac tattggctcc atctagagaa 600
gcctctcct gcggtccacg ctagatccac ccctctccac ccaccccaat ccaaccctca 660
gaatcctca tttcctcacc ctggatacg ccggccgcccc cttggacatg ccctctgagc 720
tgctcagagtc caagggcgcc ccggcctccct cttgattcag acgtctgagac 780
acaggcagca acatcagcgc ccggcccctc ccctcagcaca tcccaaggaat 840
gagccgacgaa aacacagcag ctagaagaac gcagttgctt ctagacgctg 900
cacccctcct cctctcagcc cagaggatcc tggagtattat ccggccagtt aacatcggga 960
aatgtgacag ctctgccgct ctctcagctc ccggccctct ccgggttctca aacagagaata 1020
ccagtgccca ttcctctcag gtwgggttctc cttggccggt ggggactggc cccgtggcct 1080
ccgctacca atggcagatg acgagggcgc tcggatttcg tgctcttgatt cagaaaggcc 1140
taactcgccg cctctcttac tggacttttc ctcagccactg aagtttttctc 1200
atatgaaaaa gatcccccctg tatcatcttg cacacagttcg gttgtgccttg aagagacctg 1260
agggcagctc aacaggagcc ctagacgtctg tcgaagggcg cccctccttc ccggggaggg 1320
tctcctcag cttggtctcg ctctctcttc cttgttctg acatccgacca atggcctctc 1380
catttgcttg acagttctctg cgtctgtttt ccaagtcatg aaccagggctc aagttctcctg 1440
gcgctgggg ccagccagtc ctagcagcag ctagaatcct gacacgcttt gcgctctcct 1500
cgctccttac tggataagg ccagagccgc ccgggctggc ggcagcctg aagttcttctc 1560
atttgtgtg cagacccagtc cccccctcag ccgggtcttc ccaggttcac cagccctctg 1620
ccttgaaggg ccagtcggag ctagccgctg cagccgctg acctgaggct gtagacaccttt 1680
tgacacgag gcctgatttc cttggtgatt gccagcctgc aagttgtcctg ccctctcttc 1740
aggttctctg tctactttc tctgttgccg gttggcctac ccgctcagag tcagtttctc 1800
catccttac cccagcggc gtgggtcttt gtttcttttt gttgtttctc aaggtgctcc 1860
ctctttgtgc ttcagtggttc ttcggtgctt ccggtcttttc cctagccatt aagtcagatt 1920
cattgcttgg gggatgagg ccagccagcc ccggcccttc ccgagctctg ccctctcttc 1980
cctttattgg ctagctgggt ccgggcttcg ccggccgttc cttccgcttc ccagcttctca 2040
atacagggac gacatccttt gcctgtgattt atggagcagc ccacggcctc ccctctctct 2100
acgctatgc agatacgact gcctctcttt ggtgtgtgttc tggggttctg ccctctcttc 2160
tttgttgaat cacatccttc ccagctgagc gctgctggg gacagccttc cccctttctc 2220
cattgcttgg tccacacact ccctcttccc tctcaggttc ctcagctgta aacctcagatt 2280
ctctctcttc ctagctgggt ccggcctcct ccctctcttc cctctctctc ccctctcttc 2340
aataactaca aacctggtca cctttgggaa ccaacccgccc atgtggttgc caagtagagc 2400
aagcccctct tctctctcct actacgctg cc tggcttgaga tggactttat tattagagaa 2460
gggccgatta acaactttgc agatcttctg ttggccctcga tttgggggatt tgtttggttt 2520
ttggttgtt tggagaaaaa caagttataa actgtttttt tgaatgtttgg tattatataa 2580
aaaaaaaaag aaaaaaaaag acaaaaaaaa aacaaaaaag accttttgggt atagtgcact gtgccttttag 2640
cagagccgct gcaacatctg aagacactgc acgttgagag gggctttggt gaggctttccc 2700
cctggcccgt tgaagccccc cctggttcgc gtttttggtgc tttttggcacc agaacaacctg 2760
atgggactct tggaccccttg tgtgatttcttt atgaagctgt cagggccgcc cggagccaaa 2820
cattgattct ctgtgattag tcccccctgtg tcaactacat gocctcctctg gaaagctaat 2880
gataaatac tttggcctctg cctttgggaa cctggtccac cttgtaaaac cctggggatt 2940
cctcctggtg ctgacgttacg gacaattccc atttactacag ttttggggttt ttctcttttt 3000
ccttcacattg gtaaccttaa aacaccccttg tgtcagttcg tagggttttga gatcattgct 3060
tgtatgcacg aggagcaagca cagttctagac aacctagtac cttgtgttttct ttttttttaa 3120
aatattccct atgtgctgat tattgagga aagttttttta caacacaaagt aagacctagtt 3180
gaaagttgag ctccaaagatg agaagaaagt acctgtgtgta ctttgggctgc ggtgcttgctg 3240
gtacacattg tttggtcggct ggacacactc ttcctactag aagcaaaacgac ctttttcttt 3300
atgtgaggg ggcagtctgctg aactcgtgag tggagaccaac gtgtgggtta ctgctttttt 3360
attcataagc acctccataa ttctaaactg tccccccttt gaccaacagc acaattgtga 3420
agaggtttcg aggggaaggg gggtgccccct atagaatttg aacatacagca tttctcttaa 3480
ttgagagcta attagcggcaca aagttggcta aaccggtgac tgtggcctttta attagcttttg 3540
tactaaac aagaggctcc ttttaaaccg aattttttct aatctactttg tcaaccactgt 3600
catttacatg atgcataaat taccctcgcctg tgcgtagaga tttatttggt ttaaattgctg 3660	tattatcctgt tttttttttt cttacactac ccttgggggt aatatttttttt ttttttttt 3720
tgaaagtgaat tattgctgtct tattcgatag acaataatcc tttttagcttt ggtgaccc 3780
tccctgcttg taactaatcgt gagctctggtct gtcacctgge aataggacagc ggcacactcgg 3840
tggcaatctgg attaggacgta tcgggtcccc gcgctttttc tgttttttttttt ggggtttttt 3900
tggggttttt tgtggtgttg gtttttccgaa ttttttttac tttattcagtg agaaggccctg 3960
gcctgagga gttggtgttg aatttccacac cccagcgccg cggcttctgc atctctgccaa 4020
gttgttattt attcttacacg tggtaaataa aagaatttgtgtt tttttttttttt 4071

<210> SEQ ID NO 40
<211> LENGTH: 3499
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 40
gccagaagcg ttcagagagaag aaaaatttca gaaataattg cc ttcactttgga cttggatattac 60
aatacagcgc gaacccgcttg ccaacccccc ttaaagaaaa tcgggtatatac aatactccca 120
cagaaacggc tggagttttg ttcacaaaaat cactgccttg atttttgtgct cttaatcacgt 180
aactctcttg gttttttttt gctggtgtatc ttttggggtatt gctgatattg 2400
gttccacagc acacacttgatt tctgagattt gctagaaactc tcatggagga taacacaagc 300
cctcggtgctg cagctatttt tccctgcgta atttttggaac atgttctcggt ggctgttatatt 360
cctcaatgtg atgctgctgcgtg acgtcctgtg gaagaacccgca agaattcatgt aactcaggt 420
tcactggagt tgtgctgaaga agaaacctta gtgatgagat tgtgacat cacocottaca
480
gttgaaagtct tctgctgatg cgggatgaa acaattgaat cttgtgagct tcgctgggcc
540
tctcctcaata tgtgattccc ttgggctatg tgtgatgaaa agaacgacta taataataata
600
atatagttctc cacacggca tgtctggttc gtcgctggtc tccaggagct ctgcttattta
660
ggtggtggtc tgtcttgtaat gaaacagcag cggctgcaag aaaaataagc agactctacag
720
ggagcctcat cagcagactc gctgtaagg aaaaaaggaa gaaaaacta cacacgacgag
780
ccagattcccc cagcagactc gcaaatata ttgctggaag agaagaagca agaaaggaaga
840
gggaaaccaaa ttaatttctt ttgggtttttta ctgctgaacg tgcgttcgcaaga ggtctatgtt
900
tctaaataaca tcaagctggcact ccagcagag aagaggttt ttaaatgtgct ggattctaaa
960
ggtcgctccca ggtggtgggga ggaacaaaaa acaagacctcg atatgaatt tggaggacctg
1020
ggagacgcc ttcgggctc tttctcaggg gttcttttctg ccaatactg caagccagcgc
1080
agtctgcaca agtttttaag aagctgcaca Ctottatata ataatataag tggaggagcta
1140
cacgaggctg cagcagactc tcaagctggcact ccagcagag aagaggttt ttaaatgtgct
1200
ctaaataaca ggtggtgggga ggaacaaaaa acaagacctcg atatgaatt tggaggacctg
1260
tcaagctgggt cagcagactc tcaagctggcact ccagcagag aagaggttt ttaaatgtgct
1320
tcaagctgggt cagcagactc tcaagctggcact ccagcagag aagaggttt ttaaatgtgct
1380
acgtctctag tagtcagctgct ccagcagag gagaagcagc tagaagacctg
1440
accatcgagg atgaagacatt aattcttcccc tcgcttaagtt tgaacgcctat acgggttca
1500
acccagttg caagtctct ggtgctctgctg atatcagctc ctggctctag acaactcacc
1560
acagtcggcc ttcacagctg cacaggggctg aggcatctg acgacgccca cggactctgg
1620
agtttaatat ccagctcactc tcgctctcagca caactggtga cacggtagct gcaggtcgaac
1680
atgctgctag tcggcagcag ggtgtctctct cctcaatgtg tctggtgccc tgtccaggtt
1740
cgcaggtcct ttcagctgctg tgtacgctgct tgtcagcttc ttcagctgctg tgtcagcttg
1800
tccacatctctct tcaagctggg tgcacgcttg tgtcctctttt tctctgctgct tgtcagcttg
1860
tcgtgctcactc agggcagctg cggagcagctg cgttccagcttg ttctgtgctgct gcgctgcttg
1920
tcgggagggag caactggccat tcgcttttctc agctcttcct gcgttccagt attttttcct
1980
acgggagggag caactggccat tcgcttttctc agctcttcct gcgttccagt attttttcct
2040
cagttgactg cggagcagctg cggagcagctg cgttccagcttg ttctgtgctgct gcgctgcttg
2100
aggttgaacatt aattcttcccc tcgctgaattt ttaaatgtgct gcaggttcaag ttcggtttttt
2160
ttcagctgctg tgtacgctgct tgtcagcttg ttcagctgctg tgtcagcttg
2220
cgtgtctctc tgttcggggag cagctccactt gtcgcttctt gttcttttttt ttttttatttt
2280
tatcagaagct cttggctttttt ttcagctgctg tgtacgctgct tgtcagcttg
2340
cagttgactg cggagcagctg cggagcagctg cgttccagcttg ttctgtgctgct gcgctgcttg
2400
ttcagctgctg tgtacgctgct tgtcagcttg ttcagctgctg tgtcagcttg
2460
ttcagctgctg tgtacgctgct tgtcagcttg ttcagctgctg tgtcagcttg
2520
ttcagctgctg tgtacgctgct tgtcagcttg ttcagctgctg tgtcagcttg
2580
ttcagctgctg tgtacgctgct tgtcagcttg ttcagctgctg tgtcagcttg
2640
ttcagctgctg tgtacgctgct tgtcagcttg ttcagctgctg tgtcagcttg
2700
ttcagctgctg tgtacgctgct tgtcagcttg ttcagctgctg tgtcagcttg
2760
ttcagctgctg tgtacgctgct tgtcagcttg ttcagctgctg tgtcagcttg
2820
ctatatatgct aagttatgct aataaataca ttctcaggtt atttcctaat gtaagatag 2800
tctaaatgt gcagagaaaa aagtgaacca aacaaattta ttcacttgtt atgggtaga 2940
tctgcataaa attcctcota aataaattgt ttatacact tttagccac acggcttttc 3000
ggtccattag taaactcagtg ttgtaaatg gtaacacagaa taactgtaatt ttgagaggt 3060
caggactctt tttaagggc caagaaagca acctgtgcct tgggtcatac ttggtcagta 3120
gtctgttata aagacaaata tatcttatcat ttggtcatat ttctatctg ggagggactt 3180
gggggggttt gcatacaca atatcaatata tatccacactt ttggaataag aatgttttta 3240
gattactttt ttcactcgtta aataagcata tttaatgcata cagaaaaa atgcttcctg 3300
caaaatttt agataacaag aatattttgt agatgaaaaa aacaaattatg ttgagaggtc 3360
taatgtcagt ttctctcttta ctacaggtaa ttttgtatat tttaaacatt aataatcttg 3420
aactctctaa acctttttgt atcctttttt gtttttttat taataaatct atataaata 3480
tttttaaaa aaaaaaaaa 3499

<210> SEQ ID NO: 41
<211> LENGTH: 2212
<212> ORGANISM: Homo sapiens
<400> SEQUENCE: 41

gccggtctc tgaggacagg tctgaggccc cgccttgccc gcgggccggt gcgggagg 60
gagccggcgg atggagggaa ggtggaagaat ggatactgg gaccagcaag tgccctacac 120
cctccagga aatcgcgccc ggataaagac gctggggaag cctggcctgg gcgggcgggt 180
gagctcagt gcggggcctt cctggcgcct ggtgcagcctt gaagatctct tccagatct 240
aatcggctc caggagacgt gcggcgctga aatcgcgagt cagagcagtgc agtagcagtt 300
tgctctgtat ttccattcag aaaaacctgc ttctccagcc ccccaaccca ggtacaagaa 360
ggagcccgag agctccggcc gcacagggcc cctgctcttg acctgcagac agcagggccc ctgcacccc 420
cctcccgat gcgggagcag cggcttacca cgctcagctat gaagccccca gacaatacgc 480
cattcgcctc cctggccttg atgcgcgcc cttacgctct tccccggg 540
agacacaggg aatctcggc gatcctctgg caatccocca ccccccctg gccttggtta 600
cctggggaa eatactctgg cctcagctc atggctgga aatcgcagct cttcctac 660
tcagggagg gcggggggac cccccccgca cctccaccca caatgcgctgt aggagaacctg 720
ccccacacat cccacgacga gttattaagca agatacaact gatccctgtg atgaacaggc 790
gggcagcgc gcgggggtcg aagggggggt caatggccac aggtacccag ggagggggtt 840
gtggtgtaaa cagggagcgc gcgaacctcg ctacaggca agatgtagct tcgctgctcc ctgagaaatt 900
aatgtactc ctacagaggg gtcctccttg gcctcgcctgg ggtgaggggggg ctaagcgtta 960
tcgtcattcg aacatctcg gcacactccc agatgtggct tcggtggctgc ctgagaaatt 1020
tgaagggagc atcagcaggg aaggggtcgg tgcattttag aggggaggcc ccacctcagcg 1080
cgggggtgcc cttcacggtg gosaiccttc gcgtgccctg cttgaggtc ccaacaattgc 1140
caatccactg gccctgaggg gccggggaat gtagctcaag ctcatgtctgctc ttggagagt 1200
cggccgcggtt ggggctgcc amaaagaaag gcggcctcgtc aatgcagcac agtgggccgc 1260
tcgctcgag taatcattag aagaaagcat atgggagaaa gttgggtggtg aagatcagtg 1320
gtagaaggt tgggtgctgc ccagggtggct cccccctctgt gcctctccgg gccatacagcgc 1380
tccagcttca aaggtctgag tgcagcgccgc tgccagttgcg gaggacagca ctcctttgctc 1440
ccacttggt gagacgcccg ectaectcccg agagctgggt ggcccegccg cgccacttcgg 1500
ccccaaaggt ggctaccttt actagcccccg acggctgcttt ccocctgcgc gaggctgggt 1560
cctgccctgt tcatatatatc tgaatctggt gtgggggaaa ccocctacttg aacccgac 1620
agtgcccttg ggccatggcc caagttctct caagttctcga ttcggagcagc 1680
ttcggcggct cctgagggaa gaaaggggcc gaaaagggcc gttcagagtc tcaaagcttc 1740
cctgcgggtg ttgccgctgg cccctggagga attctggtcta ctgctctcttc tgtctcactg 1800
cgccccaccc ctttccccag caacaggaaga ccaacgagagcg ttcgtgctcttg ggccgccgag 1860
agggccgttt ccacaaacct cttctgacagc cgggtgcccag gttcactctgc ccctgccagat 1920
cctccagcgc ggggagcagc aagcctgcag cccgccccgc gctggccccc tggcagcatc 1980
cggtttgcttc gcccttcttttc tccctgccgc ct gccagggct cggaggtggt aagggcagaa 2040
ccttggcgcgg gagggtgcgg gtttaagccg cagaggggctgg ccctgcctgc caacggaccc 2100
tttcgccttc tctgccccttcttaggcgc agggctgggt ttgctacctgc acctccacca 2160
cctagcgcag ccctctaaag agggcccccag ccctccccag aaaaaaa aag 2222

<210> SEQ ID NO: 42
<211> TYPE: DNA
<212> ORGANISM: Homo sapiens
<400> SEQUENCE: 42
tctgagaccc gggcggcgag ggcggcggca gggcggcggc agggcggcgg caggggaccg 60
ggcggcggg gggcggcgag cggcggcggc gaaaaggggcc gtcagagtc gtcctccggtg 120
gcgagagcc cgaggggagc aaggggcttc cctgaggcgg ggagcagggg ggttccagagtt 180
cctgccctgt cctgctctgc cccgccccgc ggagccacagc cggaggtggt aagggcagaa 240
agggcgagc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 300
ccttggcgcgg ggcttggcgc ccctccccag cgcagggatg ctcctgcccc ggtcctgccc 360
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 420
aaaggggatg cgttggcgc ccctccccag cgcagggatg ctcctgcccc ggtcctgccc 480
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 540
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 600
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 660
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 720
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 780
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 840
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 900
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 960
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 1020
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 1080
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 1140
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 1200
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 1260
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 1320
agggagcgc ccctgccgaggt ggagcagggg ggcttgcttc cctggccccc tggagggagc 1380
gaagggcaac cgcaagaaga tgaacctacca gaagatgggag cgccaggctgc gcagaatacgg
900
cgaagcgcgg gcggctcaaa aggtagaaag gaactcactac gatcagttca cgcggcgaagt
960
gctggcgccc cggggcctgg cccagagcgc ccacccgccc caaacggccc gcagcccccgc
1020
ccgggcgcgc cggggtctgc acgtggtcag agcagtagag cctcggcggc cccggggacaca
1080
ggtggagcgc tcgggccggc cggagggcgc acgtggaggg cgggctgtgc tcaccggccc
1140
cgtcccccgc ctggaagggc ccctcggcccc gcgtccgggc ctggtccagg aacccacccc
1200
ggtccgggc gcgcagttgg gcgtgagacc augggcaacc tggcaagagg cggagggggg
1260
tagggctgtgg ggaacgttca acgcgtagtg aatacagatgc tcctcttcca cccctcaca
1320
cattacatt ctcccccaca aacaagtaaa gttatacatc atcc
1364

<210> SEQ ID NO: 44
<211> LENGTH: 3094
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 44

ttctctgttta acaacaacac taatttttatt ttgaggtctg ttgatatgtg tgcataatgg
60
tagacttact atctcttctg gtgtaaaact gttcagtag cctcttggtat caagtggtac
120
tctgattcttt aagaaatttt ctaatacgga acgcatccaa tagaataatct tgaattgtgt
180
ttcataatgct gaacttgttaa ggtgggaatt aatggtgttg ttagaagggg
300
gtatgaacact gaaaggcatc atctggtgct gcatcggccgc gctttgccgta
360
acttcagcgc ttcattttgt cagaccccag gctagatgac cacagtgtgta atattgatt
420
tcataaagct tttgagttgt tggagcatct gcactgggtct gagaaactgtg tggaaggtgc
480
aacattcctac agggaggtgg aagggaaacag aagacccact gaggagggag aacttccccca
540
aacgtttaga caacagagag tagttgtcct gcagagcgcg aacctttttc
600
tccacagacag gagaagctgg cggcctttttg aagggctcga gtcgcgccgg cgcgggcgc
660
tttggctcc ggcgggttcc gcagcgccgc gtcagggcgcc ctcagagacg taaaggccag
720
ggagagggcag gggcggcag gactccagga gggccccggag ctacatcccc ccgacacacc
780
aaccgagggt ttctttgtgcg acggggtgca gcagagccct taggggtgcga gcccctcccc
840
cggagaggac ggtaagcagt gcagcgggtct gggagagcag ggggaggttt
900
gaggagttgc gcggccggcgg agggctggag agggggaggg gcgggccoag gcggccgaa
960
cggccggccgg aggccggcgc ggcgcggcgc gggcggagag gcggggtgct gcggccccgg
1020
gggggtgctgg ctctgggccc ctggctttct gcggccggtc ccgggctgtc gcggggccgg
1080
gggggtgctgg cggggtgctgg ccggggtgctgg ccggggtgctgg cggggtgctgg cggggtgctgg
1140
tctgctgggt ctctgtttct cccctctgtct cccctctgtct cccctctgtct cccctctgtct
1200
tctctctctt gatcagttca cgcgtcagag tggcagagtag ggtgagttct
1260
tccacattt cttccagcgc ctcttcagcc cttccagcga ctcagctaga
1320
tttgtgttgc cttactcgtg gggagcgatg agcagttaaga ggagcagaag tgtgagttct
1380
tctctctttt gcagagccgc gcgcggtctc gcgtgcgtctc gcgtgcgtctc gcgtgcgtctc
1440
gacatgata gcaaaaacat atcataaaaa tactgaagag cccgaggttc cggccggtct
1500
tgttctcttt cccagaggt ttcggagtta gctagagctgg tgtgagagtt ggtgagttct
1560
agtcagttgg ccggcagttgg gcggccgtaa cggccggtct cggccggtct cggccggtct
1620
gagaagataaa accaacccag ctcggtgcca caacgtctag cgcgaagtag cacatacaact ctcggactc 1680
cggcacaatt ttcttcatcct accctacaact cggcagcttct ctcacactta aagcttttca 1740
aacatatgct gcaagttgaga cccgagcag aactggcagc gaaaaaagct ccccagcagc 1800
cccacacac tgggtcttccaca ccaacagc aaa agaaacccg ccgctgtaa ccttcagctg 1860
ttggctccac aatcctcata tttctcctc actccagagaa aaccttcaag 1920
cctggagac atctgttctt ccaaaatgct cctccctgga acggccaaacc tctggacttca 1980
agaataagc tgcattgtgct accacacaca accattgctg ctatcaccct cttcaggaacc 2040
cctccgacgcc accttgaacca ccaaacgttct cccgacgaca gagcattgatt 2100
cagtgctgcc ctcgcaagag caccctgagg aagattctgc ttcgaggact aacagaacc 2160
atccagtctg aggagaaagc gacaagttta aatatttac caaatcacaag aagaaacc 2220
gttgagact gcacccacccg cctgtctgac caagctgatga ccaagactcct cctgggacta 2280
tgacagtttc atcctcataa gttttcttctt caagccttct tttcagcctt caccaacctc 2340
tgcacactct aagacacctt ccattcctgta tcaatctctgt gactttctgc agtctgttga 2400
cctccctacag ctcggaagctg ctcaacactt cttccagttt cttttcagtc 2460
tgaaacagc tggagccccct cctgtctcttg ggtgtgattg agcccctcacc cctgggccct 2520
tttccccgct ctcaagcag cttcactcagt cttctctgtt gaattgac gagaaggagac 2580
gagaaaaagc accttcaaaactg tgtttgtttct ttcacagcatt ctttatattttt 2640
tggaggtttt cggagtattt ctatattatt toaataacgg gcttatatat gcttttcccct 2700
tttgagtgct ggcacctttg cttctttttt ctatctctctc ctctttcctct 2760
cacccctcccc tgccttctc ctttcctcttt tcccttcttct tttctctttct 2820
aaaaatatttt tggatgtggt tgcagagagg gtttgcagtg tttttgtggg attttttggg 2880
caaaaacagt gaaacagctc cttcctctct gttctctgag ggacccctttt gccaagaaac 2940
attatgtta gaacctttttct cttttacataa gattttgtgag aagaaaaaaa aaaaaaaaa 3000
aaaaaa aaaaaaaa aaaaaaaa aaaaaaa aaaaa 3034

<210> SEQ ID NO 45
<211> LENGTH: 3077
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 45

ttctctgtta aacaaaacc gtaaatattt ttttgagagtt ttcgctagct gctctatattt 60
atgacttttcttctctctct gtctaaaact gtcctctctctctctac tcaagtttgac 120
cctattttttt aagaattttt ctaatcgggc actctctctctctctcc tgcacagctg 180
ttcatggcaaco tataatctttt tataactacta aacaaacttct aagaggtttt cttctctacc 240
tataaatattt gactctttttt gagaggtttta ggtgtcttaa atatgtgagttctttc 300
gctagcagca gataggttctt ctggctctct cttcctctct cccgagcagc ttttccacg 360
acctccacgct ttcctattgct ttcgctccag cttagagacg cacaagtctta attattatt 420
ttcctttatg ctggagagtctt ttcgctctct cttcctctct cttcctctct cttcctctct 480
acacacccc aaggagagtc gggtgagaggt gggagccctt ggcctctctctctctcc 540
agtttagcct ccaacagacgaaccttctctctcctctctctctct ctgcacgcacg ggcctctctct 600
tctcaacgcct gcccaagttt gcgtgcttttctt ccgctgagcag gttcctccgac ggcctctctct 660
cctggctctc gcggctctct cggacggcgc ctccaggaacc tgcagacgcc 720
ggaagggaga gggcgagac gcctcaggg gaagccgagag ctacattccg ccagccaaacc 780
aaaccgggtg tgtgtagagc cgctgctcaa cagccacccct taggccgccaa gcccctcccc 840
cggaaaatgag cggatacggc taacccgcttc gggagagccgc caggagcgct cggggggttt 900
gagaagggagc gggcgagagc ggcggctggag aggccgcccc ggtgagcgctg cgggagggga 960
cggagcggcgc gcggctgtag cggcccccacag accggcgcg cggctgacgc gcgggaggggc 1020
gggagggagc ggcgcgagagc ggcggagcgc gcggagagcg gcggagccgg gcgggaggggc 1080
ggaggctgagc tgggaggcgc gcggcttctct acgctcgccg ccagggaggtgc cgggaggggt 1140
cggcggctgc ctcgatgttc cagcggtgag agggaggtgga gcggagagag gcggatgcgtg 1200
tggagggcag aagccagggg gcggcccgcgc gcggcagcgcgc gcgtcattgc agtacagcgtg 1260
cattacccct gcggagagtc ctcgcttgcgc tctgcagcgcgg ctgcagcggcgg ggcagatgtg 1320
tctggttgggc cttcaatgat gcggagttta agcctttgcca gcggagagag gcggatgcgtg 1380
tctgggaggtg tggagaggtgc aagggacaag gcgtactaca gtaaatattgt cgaactcagca 1440
gactactggta tgcagggaaactcagaa aaatgggaatg ccaagagacg cggagaacg 1500
ttgcttctgg ctcagagatg tggagaggtgc aagggacaag gcgtactaca gtaaatattgt 1560
dactgtaagag aggctacgtct agtgacaca ctgagcggcgc gcggagagag gcggatgcgtg 1620
gggagagagc ctcgcttttc gcgcagctccag gcggagagag gcggatgcgtg 1680
cgggtgctgatta ttccttcattt actcctactt tccggactct tcccaatgtgta agtctttctca 1740
aattgttacg ccagcgagag aacgggagac gaaattaatatct ctcgaggaac 1800
cacacagct cargctcctac tttgctgaag ccctttccacaa aagacggcgcg gtcgagccct 1860
ttgctgccacc cttctcaacttg gccgaactggt ttccttcattt actcctactt tccggactct 1920
cgggtggtctgctggcgc gcggagacgcgct tcagcttcggaga gcaggcggcgc gcggagagag 1980
dactgtaagag aggctacgtct agtgacaca ctgagcggcgc gcggagagag gcggatgcgtg 2040
ctccagacca gtcctagctg ctccagtcacat gacagttgtg 2100
cagcccagct ctcgcttcccct cggaaaggtg cgaatctgtc ctcacagctgctacggt gcgttttctca 2160
attctgttct gcggagcaggc gaaatctgct gcggatgctgc aggagccgagc 2220
gttctgtatttt ggcgtgactc gggagaggtgc ctcctctggttt cggagagag gcggatgcgtg 2280
tgagccacct cggagagag gcggatgctgc aggagccgagc 2340
cttctgttctgc gcggagagag gcggatgctgc aggagccgagc 2400
tatatgataa aagttgtgctgc gcggagagag gcggatgctgc aggagccgagc 2460
tctctctctt tactcagagtttt ggattggcattt cggagagag gcggatgctgc aggagccgagc 2520
ctttttttt aaaaaaattttt attttttttttt tttttattatt gaaatttttgaa 2580
cttctctctt tactcagagtttt ggattggcattt cggagagag gcggatgctgc aggagccgagc 2640
gagagttgtgctgc gcggagagag gcggatgctgc aggagccgagc 2700
tctctctctt tactcagagtttt ggattggcattt cggagagag gcggatgctgc aggagccgagc 2760
cttctctctt tactcagagtttt ggattggcattt cggagagag gcggatgctgc aggagccgagc 2820
ctttttttt aaaaaaattttt attttttttttt tttttattatt gaaatttttgaa 2880
ctttttttt aaaaaaattttt attttttttttt tttttattatt gaaatttttgaa 2940
agtagttgctgc gcggagagag gcggatgctgc aggagccgagc 3000
tctctctctt tactcagagtttt ggattggcattt cggagagag gcggatgctgc aggagccgagc 3060

<210> SEQ ID NO 46
<211> LENGTH: 1443
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 46

gggggggggtg atagagaggg agcccgagac gccgcccaggg agacgagaccc caaggacggc 60
gagccg gcagcagcg gttgaaaatg aacgcggcgt gtagcatgt ggaaagacc a 120
gggaggtg ggggtgataca gtggcttgcag tggccctaca aaacagagtc acoccaagc 180
tccgagaga tccagctgtg gcctcttttact ctggagatgc tcacagagga aaggttcgcc 240
catgtcatcg ctgggagcag cggagagtaag cggggagtttg tcacaaagga tccagatag 300
gttggcgcgctcttgagcgc cggagagtaag cggggagtttg tcacaaagga tccagatag 360
cggcccctca gatctcttta caacagcagg atccttctca aaccacagga gagaagtttg 420
acccataaat ttaacatcatca aaccgtcgtg atgaccaact acctcttaca aaccatcgg 480
tcagttgtgata cagtaaaaaac tctttgtgta gggtaatagt ttggaattga aagatattttt 540
taaatctcca aacctgacag atgccgatgt tagaagatttt tagaaacacca aacaatatttg 600
tgatcctcgttgatggtcctca atggtctttac tcaattctca acacattacag gtaggtatt 660
agctacatca gttacagat ggttctcttctttctttcag tttctgtgatt 720
attgtgtgtg atcctcttaca tcaattattctgt ctggtcttggatt gggatagtga ttctcttattg 780
tttggtgatg gttttatactg caagactcag ctttaacacca caggccagtgt gttttcttctt 840
agatggagct cttctcttcac aaccacagac gatttctgtgt ggaactacac aagctctcctc 900
taccaagtaga ctttcttattg ctttggatgatc gagagggggg aagttcatcag 960
gttgcaagtg aaccacagtg cttccttgct atgtagagaa atcctcagag aaaaagattc 1020
tggtggtatct ctggtctctc ctggttttggtc gtaggtaggg cttccccctct tgccttttctt 1080
gcttatagc tttcatccac gcctcctcca ttctctttgt atgttgaag cagctcttga 1140
aaaagattggtg tgggtagattgt gttttatatgtc tgtttatatgc taatcattact 1200
cggcctcagc aatcattagtg atctctcttct atgtgtatttt aagactatctc 1260
gtagtttaga caccacacag acctgctcaaa tctcattcct atatattgat ttcttttcttc 1320
cottaaccc aacatcatttt atctcagtttt tataaaatt tagaaagagag a 1380
agataattgacaatattataa atagaaaaat taatatataat aaaaacaaat aaaaacaaaat 1440
aaa 1443

<210> SEQ ID NO 47
<211> LENGTH: 229
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 47

cceacagtgt accacacacca aacactgatct atatgtcgct acatatgc acatatgtcg 60
cagagacgca gcctgcgcgc gcagagacgca ctccagtctc tgcggtccgagcg cagagacgca 120
tcagagacgc gctctgacg agacgacgca ctctctggcct tgcctctgacg cagagacgca 180
gacagccgca acgcagacgct atcagctgtg ggggagccca ccggggaggt 240
cacaaggtca caggtgttct tatttcacat gcagcgtcgg gtctggaggc cctgccctct 60
cctgccctga aagctctcct ttgaggaacta gcaagcctgc cccttcaggt aagttggac 120
tgggtcagcc acacagcact ctgctcttac gcacattcct ctgcttcgcc gatgctcttg 180
actgctctgt tcgcaattga agagctcact tctgcctttg agcatcagac acgctctgac 240
tctagctgga cactcagcct cccctgtaac tgcgctagc gcgcattgtg ggagttggttc 300
cagttcctgt gccagcagta caagtggtgcc accaaattgca tctctctcttg caaccttcac 360
atcaggggct tgccagctgt ccagatcaga cggagagagt tccagctgtcc agctggctcc 420
tgcagcagact accgtcaatt catctccccag aacatcgcga ccaaggttca cctccttttt 480
aattacgctg aagaaagcga gcccacaccct aagacactgt ctgacttccca ctgctgtgaa 540
aacaagttcga tcaaaaagctg agacagcact ccagacattct gaaacgtgac ccacacat 600
cattcataaag aatttacttg ccctctttcg aaggaacattc ctctgtgatcc 660
agaaggggag ataggggaaca aggaatctttt cggctggtcttta aatacgagact cctgggcaaa 720
atgtaggcagc aagagagataa ataaggctatg aaaaagttag cagagcccttg 780
agataactct ataaacacgg aatttttgag ccgtttgacg gaaagttgac gtaaaatctt 840
ggaaaatagt cacaaggttg gcagagagact aagctatgat cggcggccag cttcagatcc 900
atttattggat ttatgtcttc ttaaaacaaat ccagaggctgaa tagacctcctg aaaggtgctca 960
tttctcttct tttttttttt aacotctgca aacatgctgtata aattttcttc acatacctacg 1020
ttctaaaaggg atacagatgt gttgctctcg gaggagagt gcaagacact ttaaagaaac 1080
cattctctct ccctaaagggg tcaaggggat agcttttcttg gttgcttctg ccaacttttat 1140
ttttctcagag cagagaagcc actgctagtg gttgataag ggacaccactctg agaagacact 1200
ttcaatcctg cccttctagaga ccatccatct gctccctctcc ccaagttcct aagattttttca 1260
aataacgtgtg gcagccctcc aagagttgggg gaaagctgtga ttagttgaca ggttatggtga 1320
ggagaggtgtg agatataaga cactacacct ttgataatttta aatattaataa gcaaaaaact 1380
catagaagag atcctcccttt cccccccc cggagaggt gtaaatctgt tcgccagggc 1440
tggctttgga ctctctatgt caagtgatac cccacacccgc ccctccaaaaa gact ttgctgtg 1500	tacaaggtgt agccacgcga cctggggcgaga aaataggctt aatttataaag agaotaagc 1560
catcaagagt gctgcagagct gcagagctgcgc aacagatcctt aacggcttgg cattttttctg 1620
aatttttttt gcccacccct gatgaggaga cagaggtagctg ctgctctctg caacatctcgt 1680
ttaaccacat tattaaggct tggctgttttt ctaccctctt caaatctgat gctggactct 1740
atatcagatc gttgctctttt gttgcttcttc aacaggtt ggtgtttttga gtttttttttt 1800
attcagagtct tcaacaaat taaatccggag tctcaggtgttg cctctcttcca aacagcttct 1960
aagcgttttag ttaaccagctt gtttgctcgg accaccagcct aggctacatgcc aatgttggtct 1920
cttttgggtc tttcctgctt acottactgtg tattcagttg ctattaaagcttatccttgg 1980
acaaaaatcg 5tgcctcagag ttttttttcttttaagagacctc acgacagagtt ggcaaggtgct 2040
catacttcca ttcatacagtgt gtagttagttc ttcaggtgagcgagcggtgagt tgggacagc 2100
cctaggacct gttgctctccac gcagggggagc taaagttggg gtgggtgtct gtaacaca 2160
gggctgaacc cttttttttt tgggtgaacct aagtttctgg ctagttgtaaagttttttttta 2220	tactttttttt ttaaaatattg gttcttctttg gttttttttttctttaagaagttttttt 2280
atggtatctga atgttatatgc agagagagatg

2319
accctggtg cccatccct atagagactg gttgagatgc agctgtgtgc ctccctccta
60
tgagcaccg ctattgatt ttcacccag aacttctag ggtaatgagat catgattctg
120
gacccagag tgcttaatgaa tcctaaacccc ggcacacacc tcctcaaca caagccgacg
180
ctgcaacgac gtaaactcag gtcaaatgtt tcctagttgg ttttgggag ccagtggtcat
240
ggaattcagc ttcgtacgct gcacaggtcag cagagcgtgag ggatgtccca gcaccttccgt
300
gcaccacacc agctgtagtc caaattgtac cttcttacga agtggagcat caagccgag
360
caactcgtg caagtggtt gcaagagcg ccgccgacgg cgagggaggg ggccgacgct
420
tcctacgaca agttgcaacgat cttgaagttg aagcgccagt gcagtaattg cacgcttcag
480
tcctacaca atgtcattgt gaagactgaa caaactgagc ttctcatcat gaaacactgg
540
aagaagcagaga aatcatattta tgaacacacat ttaggtgacga ctgtaaagtttt gttggaacgc
600
aaaaacttctt ggcgggctcag gatctattga ctcaacaccca gtcacaccctc cttgcgtacag
660
tcactctgata tgaaacaggg gcacctagcc cctgcaacgct gccacacaca agaaccaccac
720
cgcagagggga ctcctagtac ggataattc cggcacaatcc tctggaaccc agaacaagac
780
ccaagattaa taaatggga aagaagattct ggggcccct tcaagctttt caaatcagacg
840
gcagctgctc agcttagtggt taaaaaggag aacacacgag cgtcctgct tgaatatgcg
900
aggcgagcata tgaatatatg ctaaaagaga gaaacctttta agcggttgga ggccagaaga
960
tggtatata aattttacgg gattgccgag aaaaagaaa tgtggagtc gc
1020
cataaactt tgaccacacc aaaaaaccaca ccaattactc agaaccacag aaacctctgyga
1080
cgttaaatatt ttgcacctct ttttatttt gatttttttt acatattttt gttttattcg
1140
cattcttact cgggagaagg aacaactata gtcgatttta aaaaattatttt ttttactctg
1200
aagtattcag ttattgggga aaaaaagtac caaattttttttt gttactattg agtctgttctg
1260
tggttctgt atttttttttccc tgttttccttttt actttttttttttt cagcggagatg caaacagtat
1320	tgttagcctt gttgctttttt ctaagagaaga gaaaaacaaata acgaggcctg attaattgctg
1380	tgttagctga catgtgtagg aaaaaatttg tcgcttgcct ttcacataagc acctattggg
1440
ttgtgcttca gggagttgag atgtagttcg cttaaatattt cgggtgctct gaagggatttg
1500
aacaagttgg ggagtctcgg gaaatttaac ctgagcctta ggaaccaatg agtgtagata
1560
ggcgacccct aaaaaccacaag gtgggattata atctctgctcat aaccacattga cttggatcgat
1620	ttaaaactag aagaagatatt atgggaggggg gttttttcttttttt ctttaccacatt ggattttcctg
1690
atgccccggc tgtcttcttt ttcctccctt gctttgctgg cgtaatactt taatttcctctt
1740
gttgccctc tccttaggtta aactagagaga acctagctgg gctttttttttt gttccactag
1800
cctaaaatta agttggaaaaa aaggacaggg ggcttgaggaa agctgagaa gttttccttt
1860
gtctttaaact gtcacagaggt ttcttacacg tcaactatct gattttttgggg cggagttgcc
1920	ttggggctga ggtctgctttttt gattgacagct cagttcttct ccttttttct
1980
taatgcttc acaccaagaa acacccagct gttgtaacgga aaaaagctgg agaattttttttc
2040
cataagtctc ccacacccagct cttgctgagc aacacgagttttt ttagctgttc
2100
aatagagag tagtyggttta ttctcaagag aaaaaagat cttaggttttt ttgctctttt
2160
agagtgaac aaagaaga aa cttcaacctc accctatctt tattttttta aaggtagggga 2220
cataagagt aatactgctg aggagctctt tttcagaccc ccagatgaga gcacactgca 2280
gataaagta gcaatagtc agtacgaa aa ctaaaataga agacattttc acctgaatatta 2340
caaagcca aa ttaaatatt attgtagaag gaaacaccaaa aaagaaatatt cccagggaaa 2400
atctctttt caggtatttta ttttaattat ttttctttctt ttgatattac ttttattcgtg 2460
tctcatctgaa catagcaggt atgttgcttt atttgcatag aggagaaattttt 2520
ltttttttttt tttttttctt ttaaaagaaag aaagaaagaa aaaaaataatt aacagaaaaa 2580
ctcagctcag gcacatgttt caccaagagct ttaaaagctttt ctctctcaat aggagaaatgg 2640
lctgaggggt ctctgagact tataattgag ccaatcaact ccggagcaac aagaaagaat 2700
agacacttcttg aaatagttatt tgaatgctctt catctccttttt gcagatcttttttttggc 2760
tctctgctcc tgggctttgt ccccttttct acctctctttt cttctcaaat aatgacaaca 2820
tgagaatag ttcatacgaa aagatgctgtt cttctcataa tataagagc ccatacagag 2880
tgctcatact cggtagaggtt agcaaatctat atctcaaggtt ttgcaagtttg ttgcatataa 2940
tccagaggg ttggagaaag aagaaataaa aattttctctt aatctcaatag aagacaccttcc 3000
taggtgatga gtctttgagc caacagataata cttcttctgaa ggtgcacaccca aatctctatttt 3060	
tatcttcctg cttgctttgtc acctctttgaa accttttaattgtgctttt aagttaacatg 3120
lactggaaga gttgagaaaga aaccaactga acgacaagcc cttctgactgt gcacactcaga 3180
lcatctcgca gaaagaaacat atgttgctttt aacaagagcc caacagctaat ttgaggcttcc 3240
cccagcattt acatctgttc acaatagtaat aataaataat caggtctacc gatgggtatg 3300
aatgagactaa cttactggttctt tctgctggttg cctgctgtatt tattgggctac 3360
lctgctcataa cttagttcgtact gttctctctt cagctctcttgctcagcacatcaatataaagggagg 3420
tcgcatactt cgtgacttct caagagctcc acaatctgtaaa gtcttctcat tctctcactaat 3480
tgaggcggc tctttttctca aagcctctaa tcaagagcnaa ttggctgaatgg agatagcttg 3540
aattggagga gtcaacaactt caaacaatttcc acagataatttt attggatttgtgtgctttt 3600
agasagagag agatgggaat 3617

<210> SEQ ID NO: 50
<211> LENGTH: 1894
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 50

gtctggacctt ctcgccagac atccctgtcag aatctggcgtg ctctgctcggct cccagaccc 60
agctcctcag gcctgctgcttg gctctgctcag gcagccctcctgc gttgggctgtg tggcaggttg 120
ccagccggga gcctggtggttg gggtgtaggag gctccctactc acgcctggcag ccctgcagag 190
tcagagggcag acocctgtcag cctgggctggcc cccctctgttg ccacactgga tctgctgctct 240
gccacagcct cttcgagcgtc cttctcacgg ttggccccca tcgctccgggc tggagacagc 300
cagttgccctc agtcgctggcag cctcttttcctgcgccctgaa tgggcagactgc agcagacagc 360
ctcctgaggg ccacagcgaggt taaagcaacac agcgcgacg ccacacccaggg cggccctgggg 420
cagggccgcc cggggtctgct ccccgagccctc cccccgcccc ggggggaggc 480
ggtgtgctgg aagcaggttggt cgggggtttgg ccagagctag ccgagagcagagcgggag 540
tctggctgcc agtcggccacct gcaacgctgcc gcggggcctgt gcgggtctctctt acctctccctaa 600
cattggactg cttgacagctg ggagggcgcct ccgggagctg cggggtctggc aatctcctcct 660
tgcggaggcg ccacctgagc agcctagcca gtagccggtc atgacagcc aagccccagc 720
gggcagctg gcctgtggtc ccggcgggct gcctgggag gcagctcggc tggagcggtg 780
gcagctcag tggggccgag aagttgctaa gacacattgca acgcctcgtc aagttcctaa 840
cacacccgca gatccttggtc acggtgctgc gcagctggtg ggtgcagtctg 900
agagacccaa tacccgctgg ccgctctctg cggagcggct ccgagccagga 960
agcttgctgcc atgcctggag agcagctcgc ccgctgttgc cccctgggct ggtgagttgct 1020
gcagccgccc cttgagctgt ggaagctcgc gcctctggtag aagagccgagg ctttacctgg 1080
ggggattccc taatttgtgct cgaaccgtga gggaggtgag acgcagagcg aggtaggacct 1140
atcatgctcc ggccagcoca ttcacctgtg gcagttcctc aaggtggtc tatactaaagc 1200
ccacagatatt ggcgcttcca ttggggtgct ccaacagaag aagggtcatct taaaaaatga 1260
ggacccagc cagttggcgc ggcgtggtggc cattgccgac aacgctcccg cctgcaacta 1320
cgacagagtg agcgccttca tccgacagta ttaaacaagg ggcagcttcc gcagagacag 1380
caccctccac ggttgctgtc ttctgcttca ttaaagttcag cgccgctttc ggggaaaggc 1440
tgaaaccggct cccttgggct ctgctttctg cttggcttctc tccacctcgg ceccctgatg 1500
ggggaaaggc gcgcgtgctgc tggctgcctc ttcactctcca cagcggcaagg tccagggaggg 1560
gccacacact cccccggggt gattgggctc ctcgctgggct cccggaggcc gtcggggggg 1620
gtgctttccc cttccggaag gcctcttcccc tggagccagc aagggacagc ggctgtctcc 1680
cacacacctgc ctcctgaaccgc agcatttcgc aaccggagcc tacaagggc cagtgaactg 1740
acaagggcgc cagccgacgc agccttcctg ctgctccata ccccttgcctc ccactctgca 1800
ccacactctg cattggctgc cggagcgctc gcctccctgga gtgggccagc caggagttgc 1860
cgcgggaagtc cattaaagct atactagagata ctg 1894

<210> SEQ ID NO: 51
<211> LENGTH: 2180
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 51

gggcgcagaa gctgtttgac acagactgca caccgcctgg ggaataagc agtaagaagaa 60
gtgaacgcgg tcggctgctg tcgctcaacc tctgtccctc acacacaoca gcgcggaaaaa 120
aaoagagaga gcgcagagaa gaaagagag gggaagaatg agatctccat tacaagagcc 180
acacagcttg gcacacgcgc gtcgacagtgc gaagttgctgg gcgttgcctc gcaggtcctc 240
tagaaattccc cccccggagata cctcgggagtag gcagagatgc aatcaagctgtg 300
tgcgacctg tggctggctt gcagaaaatg agcatttgat cctgggtacc 360
tgaaagctgtg gtgaacaggtc gcttcctcaa gcggagaagtt cttggacgatc 420
cggaaaaaacc ccaaatagtat aatcaatgat aagctgagca gcgctctggc atactatat 480
gacacagaac taacttacagag gttgacggag cagaagtttt ttgtcacttt tctctttc 540
ccggagatcgc tgcagattgg cccctcagg cggagacttc gcggggagag gctttctgtg 600
cggacgagc acctgcgcgg gcgctggcag ggcgcggcgg cccgacgaaac ccggctggtgc 660
gctcctggac gccgcagctg caccgataac attcacttcg gcgtctgcc tctcctccac 720
attactctc cggcagacgcc cccgagcgcgc cgggtggagag gaaggctggag 780
gggcgcgcgc gaagagacgcc cccggtggga gagaatgagga cttgtgagcc 840
aataaaacc gcacagcagct caccgagcgg gttggtgtccc gctctcccac gttcaggggt 900
gcccagggcct ggctcctcct gcgcctgccc gctgcccagct tcctcgcttct tttacttg 960
catgcaggt gcagagactta cgcccttcgct caagacgcttc gctgctgtggt ctctgctcctg 1020
cctgctcctc tgcgttcatac cctgctgcct gcctgctgctgc cttggtggtggt cgctgct 1080
tggagggaggt tggagggaggt tggagggaggt tggagggaggt tggagggaggt tggagggaggt 1140
cgggccaggaggc ggccccccc ggggggggggg ggggggggggg ggggggggggg ggggggggggg 1200
ggggaggggg ggggaggggg ggggaggggg ggggaggggg ggggaggggg ggggaggggg 1260
ggggaggggg ggggaggggg ggggaggggg ggggaggggg ggggaggggg ggggaggggg 1320
tggagggaggt tggagggaggt tggagggaggt tggagggaggt tggagggaggt tggagggaggt 1380
cgggccaggaggc ggccccccc ggggggggggg ggggggggggg ggggggggggg ggggggggggg 1440
ggggaggggg ggggaggggg ggggaggggg ggggaggggg ggggaggggg ggggaggggg 1500
tggagggaggt tggagggaggt tggagggaggt tggagggaggt tggagggaggt tggagggaggt 1560
gttggtgctg cgccgctcgt tcgcctgcctgc gcgcctgcctgc gcgcctgcctgc gcgcctgcctgc 1620
gtttgccgctgc gcgcctgcctgc gcgcctgcctgc gcgcctgcctgc gcgcctgcctgc gcgcctgcctgc 1680
cgggggggggg ggggggggggg ggggggggggg ggggggggggg ggggggggggg ggggggggggg 1740
cgggggggggg ggggggggggg ggggggggggg ggggggggggg ggggggggggg ggggggggggg 1800
gttgggctgg ggggggggggg ggggggggggg ggggggggggg ggggggggggg ggggggggggg 1860
ggggggggggg ggggggggggg ggggggggggg ggggggggggg ggggggggggg ggggggggggg 1920
tttggtgctg cgccgctcgt tcgcctgcctgc gcgcctgcctgc gcgcctgcctgc gcgcctgcctgc 1980
ggggggggggg ggggggggggg ggggggggggg ggggggggggg ggggggggggg ggggggggggg 2040
tttggtgctg cgccgctcgt tcgcctgcctgc gcgcctgcctgc gcgcctgcctgc gcgcctgcctgc 2100
gttgggctgg ggggggggggg ggggggggggg ggggggggggg ggggggggggg ggggggggggg 2160

<210> SEQ ID NO: 52
<211> LENGTH: 3171
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 52

| GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG |
| GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG |
| GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG |
| GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG |
| GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG |
| GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG GGGGGGGGGGG |

2180
ctgtggatt caaaggctgt ctctagcctt tgggaaaaag ataagaacaa accagocatg 900
aacatgaa acatatggag acgcttggag tctactacc aaaaaggaat tcttgcaagg 960
gttggggagc agagggctgt ataccagcct aaggtataac gaaaaaacc acgtgctcata 1020
gatgagaca aaaaacaggct tctgtaatgg aagttgcagc cagactaagt tgaaaaatca 1080
ttgaacagag ggctccagct tcgaaaaagtt ctcctgcaag cagataccct tgggtcgagt 1140
gggaaaaatt caatccctat aacgtcgtcc agagcagaga aggtgtgata tagaggttgtg 1200
aatatcactt cccctgggca cgtgctccaa tccaggtcttc ctcatacaac tggctcctg 1260
tcagcaacag cagctcgaag gacagccgtg gttggmatgg aaggtacctgt tgtaattcag 1320
tcaatggtgc agaaaatttc aacgtgggca gttcagtccag ttaagcgag tcgccacatta 1380
ataacgacga ctatcgcac accacccggc tcgtccaaaggc tgttcacctc gacataccct 1440
acgtgatcg cagctcctac tcgaaattga gcacaaaaata ccagctgacgc tgcgaanaat 1500
attaccatcc cagcatacaag gtttgacag tcgctactag aagaaacgtt aatcctgatg 1560
ggctagagga gcaattcata tgggtgaaacc ccattggtctg tgaagacact taccctggtt 1620
tcaatagcctt atggcataacc tctatgtgag ctacatacgc ggtcctgccc 1680
cagactcctc cctcagtttac cagtgcagtct ataaaggggg cagaggttaaa atcggagcagaa 1740
gttgggggaa gcagcagaca gttggtgaaaacct cttgtgcaag tagctagaga aaaaaccgca 1800
gacgtgaaata agagacgctac ccaagttagtt gttgctagtg ccgctccagc tatgccccct 1860
ccctgttaacta tgaaaacagg aaggtgtagtg acagtgtaga aataaaatag cagctccacc 1920
agtgcctcctgc cgctggttagt aggcaacttg acataaaacat tggcaagggga agtcctcagac 1980
aaaagtcacaa gagaattctta aacacattttt atgtcataaa cacaacaacat cagactttact 2040
ggaaataaat tacatattgct atgtttcgagt gggaaattgaa tacatatattgag atgttgtagc 2100
agaaaaactgc cttctacagt agggacacac tcagcctcct ataaagaaaa acgtggcagaa 2160
gggaccaagc agctcactac gatacatag tccactcagc tgggaacac acataccctg 2220
taagaggtta taggggtcag atgggggggt gttgggtattg gtaatccac gcctacatat 2280
agcataattt ggtgatcatt ttagcataac cagcattatat ttcggtggttc gcacggttctt 2340
cacataaagcttgtatcatg aagacctgct atatataggt atttttagat gttgaattgaa 2400
gaatgatcgc aaaaacttgg cgaagaagaga tacccaaaaa aacctgatgg cgatttttata 2460	tatatatata tatatatata cacaatatca tatatatcat atatagctaa ctgattaaaaa 2520
acaagactt tagacaattt ttcagattttc tttctggaaa taagtctggat gttggtttgtt 2590
gtaaaaaaattt gggaaaaatta aaaaaattgc tccatgacac aggattttta 2640
tatatatata atatctaatc ttagattcgt atgcctttttc gacagcttttacat 2700
cataattcag ctgtgggtgtg gctctttttgt gcattggcgc tgggctcactt atggggcctt 2760
gggctgtttcg cagactacac atttaccagtt ctggctctac cgggaatctt aggctacgcc 2820
tgaatatcttg gcagaaagtt cttataacgtg gattttttct tgggtttttttttt 2880
agtgattaat ttagaagga ataataaggttg tctgtttttgg agctatttttttttttt 2940
cacaatttgc agtatactct attttgttgg attctttggtt caaagttgtat toaataggttt 3000
ttggctaat ataatagtaaa aatattctttt ctggtctttgt cttgggtatg attttaaaaa 3060
atatattaaa gttggtagatt aagttttttta gtttaggaag tggtaaacc ctctctctctc 3120
cctccccacac acctgattgaa gaatacttact tccctattata ttaatatgagg 3171
aaataatgta aggtgtgcta gactacttta cataacaaact gttttcttgt taacatcatt 60
tagaaagactg gatttctggta taatctcttc actcctactc tattgacctt taaaacatga 120
taatgacaac ctataaactt ggaaccacct agtnaacctt taatctcatt gattaatgc 180
gttgaagct toctcagggga ataacaatga catcaagcagt ggttgcagct gggaggtcct 240
ttggaggt ttcacgcaaat ggagtagaaa atcagaggg aagtgaaaaat gttctgtgat 300
tatcaagcagt gatgtcggag ccagtctccaa gtgcagagtt agagcagggc tattgcagcc 360
aggtctctgt ttatgatagtt gacactttta tggtagcaaga tggtagcaaga ggcaccaaggg 420
ttgagcaaga agatgaggaag acagtggaag cctgctctca cgcagatatt ggcaccaagga 480
cagataagac aatgtgagct gtgaagcgcct gttctctcat ggaatcctct aacgctgta 540
gggataacag aaggtctcctg ttcacctcag tctgctagag ggcagtagtc attacagaaaa 600
cgtgagtctg gggtctactc gaaagactgt aacccagcgg ttcocctctcctt attccacat 660
ccacagatag ccctgaacca atgaaaaaga aaaaagttgg ccgtaaaccct cagaccaaggg 720
aatcaccacat ttcacagttgc ttctctgatt gttatattaa gagaacagggc agaagaaggg 780
aaaggaaccc aaccatttttg tggaggtttc tttttagatct aacttaagat aaaaaaatctt 840
gttccagatg tattaaatag gattcagagag aaaaagatctg atctcagctg tggagcatca 900
aggtctctct taaagttggct gggaacatatt cagacacaaacc cagacatgga tgtacaaaaa 960
tggagcagctt ttggagcatc tacactacaa gggaaattct tgtgaaggtt ggaagcagca 1020
ggtgctgtat ttcagtcctag gattcagcga aaaaacatgt gttcagtagt gattacaaaaa 1080
gttcagcttact ttagcgggaa ctactgatga aaaaattttta gaaacggtgt 1140
caactgtcgc agaasaactcg ttcgaacagct catcttctgtgt ctggaagttg aaaaattttct 1200
cocctattaa ttctgcaaca gcagagaagg tgtagtgatag atcactcccc 1260
tgaggacgc cttcatcgcccc aagttctctta ttcagctagt gcggtagccag 1320
cctactagag acggtgctgtag cagatgacatt cagctgtgat ccacccagca 1380
aaatctccac tgcgctgcagtt cgcagctgatttg cccataaattt accacggcact 1440
gtccaccaac ccgagcctct ccacagggatt ttctgccagct cactcccact tgcggcagtt 1500
cctcctatgt aaatggagcat caaatcttca cgcagctgtag ccaatttttatt accatcccaag 1560
ttcagactgt tggcagctgt caaatgcaga ccaatcttta tctgactgga ttcggagca 1620
tttacacttgt tgcagagccca tgggctgtgtg ggcagactttt ccgcttttca atagccactg 1680
gtacacctgta aatggagcatc tcaatgctta tcagctgccc atcggcagct acctctctcc 1740
gagttactag tgcagctgtct ttcagggcag agtttaaaattt ggaaggctgt gcggaaaagc 1800
aaggtctgta ttgagaaact ccagttgatag tagagaaaaa ccagagagag ggaattaaga 1860
cagtggccac cttgagagggt tgtggccgagtt cttgctcatc tgcgcttcct gtaactattg 1920
aacaagaggg aactctgagc tgggaagaat aataatcagc ctcagactat gatctcaggg 1980
tgtttagggc aagttctctta cgcgtgtgtcact ccagaaagaa gtagaaaaaa 2040
gacatttata ccatttactgg ccagatcataa aacaacatcg aacatcagta aataataact 2100
tcactccatt ttcagctgagg aatgaactca catttgagta tgcgtgacaga aacaactctcc 2160
ttcagacttg aatcactacta aaccattatt aagaaaaaggg atcgaagggg accacgaccg 2220
tcactagct atcaagttac actaagactt ggacacactaa cattcctgtaa gaggcttatat
2280
agttttctag gggagggtgt ggataggtta atctctatgt tgtatatagc aatvtttttat
2340
gctattttta tgtcatcccag caatttattc tgtggtgctga cagtttctca ctaactctggg
2400
cctgtttaga aatctctgaat tttatatttt aatgtattct aatattatag
2460
aaacctcatg aagaggtcat caaaaaaga tactaatgctga ttatatattc atatatatat
2520
atatatat tatatatatat atatatata tattttatgct attttaacaa aagggcttag
2580
atctatcctt cttctctcttt cttgaaatct gtaatctgctt tttttttttc aaaggttctat
2640
attaaagaa aacattttaa acctctgtac ctaagcactg attgtttatg aatactacatg
2700
taatattttata taggttagtt taattctgtgt atgctctttta ttgtaggaaaa attaacagt
2760
ggctctgggc tttgtgtgta acatgtctgt gctacacact ttagctttgcg aatagtctacg
2820
acatacactttt caagctctcg tctgctactg gaacgtctgg ctcagcactg atttttgcga
2880
ggtagccata atacctgcag ttctctctct ctcttcttttct ctcttttttt gtaagtttat
2940
gagggacaata ccaggtctca gtttggacat ataataggtt gtaatatcata catttgaagt
3000
atatctatt ttttgattct ctctattaatg caatggtttt ctggaattata
3060
ggaatgaaa atctctttctct ttggtttttctt tctggaatata ttataaatgct
3120
agatagacag tttttattgt aggagagaag aaaaacctct ttttcctctct cccactacta
3180
gcaatagaa actactactc cattatatta atattttgg
3240

<210> SEQ ID NO: 54
<211> LENGTH: 1901
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 54

gcgcgcgagt ggcggcgcgc gcgcgctcct aaggggtctta gggagggggtc tgtggtgtgct
60
ggggtccgag gtggggcgcc gtaacccagc cccatoecctg tttgcaagaga gtcaggagggtc
120
cggaaaaagc aagcgctctca gcggccctacg ggcagcagctg gggccggccc atccccccggc
180
cgctccccc gtcgctgggg gggcgggggtc ggggctgggt cggggctgag ggggggcgggg
240
tggaagcttt atacggctat gcagagattg aaggggctgg ggggggctcct ctggaaggggca
300
gggcagcgtc ttataaaaaa aagggagattag cgccgccccgg ggcgcggggtc gggggggtgt
360
tccagccggc cccttgcctta gcagctggccc cccggcacacc cccggccacgc
420
cgcggctcgc ccccttcgcccc gcgcgccggg gggccgttagc cagcggccggg gggccggggtc
480
cgcggccgct ttcggccgct tgggtcggttc gcagggcgc gaaggttcctct gcctttgctct
540
tccgtccccc ccccttcgcct cagcgcccaat ctcggccccg cccggccggc cccggccggc
600
tccggccccc ttggcagcct gccggtgcctt ggcgggcttt gcggggcggg cggggggggtc
660
gacggagca aagccaggtgt gggccgcggt gcgggcccggg tttcagaaaaa caagcgccggc
720
gagcggatgtg cggcgggtct ttcgggatcc ggcgggagct gcgggggatc gcggggtgtc
780
gcggggggtg cggggggtgc gggcggatcc gggcggagct gcggggggtc gcggggtgtc
840
tggggccgc ccgggcccga cccgggacacg aactcgacga aacctgagcct cggcggtgcc
900
tactagtgct ctcacagacat ccagacgcag tgggtggtgc tggggtggtgc gggggtggtg
960
gctccggcct gcgggtcctcgg gcgggtcctcg gcgggtcctcg gcgggtcctcg gcgggtcctcg
1020
gcggctggtg ccggctggtg ccggctggtg ccggctggtg ccggctggtg ccggctggtg
1080
tctgggggggg cggggttggccc gggtggtggc cgggtggtggc cgggtggtggc cgggtggtggc
1140
<210> SEQ ID NO 55
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 55

tagccgcag ctaagcagga g 21

<210> SEQ ID NO 56
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 56

gtaggaacac tcaaaaaaeg actgg 25

<210> SEQ ID NO 57
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 57

cgcagctaa gcggagggc 19

<210> SEQ ID NO 58
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 58

cagggcatga aaagccaaac tt 22

<210> SEQ ID NO 59
<211> LENGTH: 49
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 59

aaaaaaaaaa aaaaaaaaa aaaaaaaaaa tttagcagauu gcggccccg 49
<210> ORGANISM: Homo sapiens
<400> SEQUENCE: 60

aaaaaaaaaaaaaaaaaaaaaaaaaa tttauccgac c ugau u gu c 51

<210> SEQ ID NO 61
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 61

gagggcgag ggcggggg ggc ggc 23

<210> SEQ ID NO 62
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 62

cctatcatta tctagtagtt ggata acag c 31

<210> SEQ ID NO 63
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 63

aatattaatag gactcactat aggaga aac tttcagcctg ata 43

<210> SEQ ID NO 64
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 64

aatattaatag gactcactat aggaga actc t gtagtcat tgttgt gct t 51

<210> SEQ ID NO 65
<211> LENGTH: 47
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 65

aatattaatag gactcactat aggagagca cactca aca acgactg 47

<210> SEQ ID NO 66
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 66

gcgcgggac gc c gguauccu gac 23

<210> SEQ ID NO 67
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 67

gcgg u uac c uccaggg u ccugac 26

<210> SEQ ID NO 68
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 68
gagcgccgca ggaagcuccua ucaguug 27

<210> SEQ ID NO 69
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 69
gagcgccgca gguauuuuca ggaucuuu 28

<210> SEQ ID NO 70
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 70
cgcgcgccga agcctta 17

<210> SEQ ID NO 71
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 71
tcgctaggca cactcaacac ac 22

<210> SEQ ID NO 72
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 72
cagtttgtag tgaaggacc 18

<210> SEQ ID NO 73
<211> LENGTH: 72
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 73
ggggtgcaag ttatatattc ccaaatatt ccgatatctt tcgaatcag caagaatta 60
caaatgacc ag 72

<210> SEQ ID NO 74
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 74
Met Ala Ser Thr Ile Lys Glu Ala Leu Ser Val Val Ser Glu Asp Gln 1 5 10 15
Ser Leu Phe Glu Cys Ala Tyr Gly Thr Pro His Leu Ala Lys Thr Glu 20 25 30
Met Thr Ala Ser Ser Ser Ser Asp 35 40

<210> SEQ ID NO 75
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 75
Ser Ser Ser Ser Asp
  1  5

<210> SEQ ID NO 76
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 76
Tyr Gly Gln Thr Ser Lys Met Ser Pro Arg Val Pro Gln Gln Asp Trp
  1  5  10  15
Leu Ser Gln Pro Pro Ala Arg Val Thr Ile Lys Met Glu Cys Aen Pro
  20  25  30
Ser Gln Val Aen Gly Ser Arg Aen
  35  40

<210> SEQ ID NO 77
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 77
Val Pro Gln Gln Asp Trp Leu Ser Gln Pro Pro Ala Arg
  1  5  10

<210> SEQ ID NO 78
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 78
Val Pro Gln Gln Asp Trp Leu Ser Gln Pro Pro Ala Arg
  1  5  10

<210> SEQ ID NO 79
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 79
Val Pro Gln Gln Asp Trp Leu Ser Gln Pro Pro Ala Arg
  1  5  10

<210> SEQ ID NO 80
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 80
Val Pro Gln Gln Asp Trp Leu Ser Gln Pro Pro Ala Arg
  1  5  10

<210> SEQ ID NO 81
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 81
Tyr Gly Gln Thr Ser Lys Met Ser Val Pro Gln Gln Asp Trp Leu Ser
  1  5  10  15
Gln Pro Pro Ala Arg
  20
Val Pro Gln Gln Asp Trp Leu Ser Gln Pro Pro Ala Arg
1  5  10

Val Pro Gln Gln Asp Trp Leu Ser Gln Pro Pro Ala Arg
1  5  10

Val Pro Gln Gln Asp Trp Leu Ser Gln Pro Pro Ala Arg
1  5  10

Val Pro Gln Gln Asp Trp Leu Ser Gln Pro Pro Ala Arg
1  5  10

Val Pro Gln Gln Asp Trp Leu Ser Gln Pro Pro Ala Arg
1  5  10

Asn Pro Ser Gln Val Asn Gly Ser Arg
20  25

Val Pro Gln Gln Asp Trp Leu Ser Gln Pro Pro Ala Arg
1  5  10

Val Pro Gln Gln Asp Trp Leu Ser Gln Pro Pro Ala Arg
1  5  10

Val Pro Gln Gln Asp Trp Leu Ser Gln Pro Pro Ala Arg
1  5  10

Ser Pro Asp Glu Cys Ser Val Ala Lys Gly Gly Lys Met Val Gly Ser
1  5  10  15
Pro Asp Thr Val Gly Met Asn Tyr Gly Ser Tyr Met Glu Glu Lys His
20    25    30
Met Pro Pro Pro Asn Met Thr Thr
35    40

<210> SEQ ID NO 89
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 89

His Met Pro Pro Pro Asn Met Thr Thr
1    5

<210> SEQ ID NO 90
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 90

His Met Pro Pro Pro Asn Met Thr Thr
1    5

<210> SEQ ID NO 91
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 91

His Met Pro Pro Pro Asn Met Thr Thr
1    5

<210> SEQ ID NO 92
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 92

Asn Tyr Gly Ser Tyr Met Glu Glu Lys His Met Pro
1    5    10

<210> SEQ ID NO 93
<211> LENGTH: 26
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 93

Met Val Gly Ser Pro Asp Thr Val Gly Met Asn Tyr Gly Ser Tyr Met
1    5    10    15

Glu Glu Lys His Met Pro Pro Pro Asn Met Thr Thr
20    25

<210> SEQ ID NO 94
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 94

His Met Pro Pro Pro Asn Met Thr Thr
1    5

<210> SEQ ID NO 95
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 95

His Met Pro Pro Pro Asn Met Thr Thr
1  5

<210> SEQ ID NO 96
<211> LENGTH: 28
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 96

Met Val Gly Ser Pro Asp Thr Val Gly Met Asn Tyr Gly Ser Tyr Met
1  5  10  15

Glu Glu Lys His Met Pro Pro Pro Asn Met Thr Thr
20  25

<210> SEQ ID NO 97
<211> LENGTH: 28
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 97

Met Val Gly Ser Pro Asp Thr Val Gly Met Asn Tyr Gly Ser Tyr Met
1  5  10  15

Glu Glu Lys His Met Pro Pro Pro Asn Met Thr Thr
20  25

<210> SEQ ID NO 98
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 98

Asn Glu Arg Arg Val Ile Val Pro Ala Asp Pro Thr Leu Trp Ser Thr
1  5  10  15

Asp His Val Arg Gln Trp Leu Glu Trp Ala Val Lys Glu Tyr Gly Leu
20  25  30

Pro Asp Val Asn Ile Leu Leu Phe
35  40

<210> SEQ ID NO 99
<211> LENGTH: 39
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 99

Asn Glu Arg Val Ile Val Pro Ala Asp Pro Thr Leu Trp Ser Thr Asp
1  5  10  15

His Val Arg Gln Trp Leu Glu Trp Ala Val Lys Glu Tyr Gly Leu Pro
20  25  30

Asp Val Asn Ile Leu Leu Phe
35

<210> SEQ ID NO 100
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 100

Asn Glu Arg Glu Tyr Gly Leu Pro Asp Val Asn Ile Leu Leu Phe
1  5  10  15
<210> SEQ ID NO 101
<211> LENGTH: 3
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 101

Asn Glu Arg

1

<210> SEQ ID NO 102
<211> LENGTH: 27
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 102

Asn Glu Arg Val Ile Val Pro Ala Asp Pro Thr Leu Trp Ser Thr Asp
1  5 10 15
His Val Arg Gln Trp Leu Glu Trp Ala Val Lys
20  25

<210> SEQ ID NO 103
<211> LENGTH: 32
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 103

Asn Glu Arg Arg Val Ile Val Pro Ala Asp Pro Thr Leu Trp Ser Thr
1  5 10 15
Amp His Val Arg Glu Tyr Gly Leu Pro Asp Val Asn Ile Leu Leu Phe
20  25 30

<210> SEQ ID NO 104
<211> LENGTH: 39
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 104

Asn Glu Arg Val Ile Val Pro Ala Asp Pro Thr Leu Trp Ser Thr Asp
1  5 10 15
His Val Arg Gln Trp Leu Glu Trp Ala Val Lys Glu Tyr Gly Leu Pro
20  25 30
Amp Val Asn Ile Leu Leu Phe
35

<210> SEQ ID NO 105
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 105

Asn Glu Arg Arg Val Ile Val Pro Ala Asp Pro Thr Leu Trp Ser Thr
1  5 10 15
Amp His Val Arg Gln Trp Leu Glu Trp Ala Val Lys Glu Tyr Gly Leu
20  25 30
Pro Asp Val Asn Ile Leu Leu Phe
35  40

<210> SEQ ID NO 106
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 106
Asp His Val Arg
20

<G1> SEQ ID NO 107
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 107

Glu Tyr Gly Leu Pro Asp Val Ala Leu Leu Leu Phe
1 5 10

<G1> SEQ ID NO 108
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 108

Gln Asn Ile Asp Gly Lys Leu Cys Lys Met Thr Lys Asp Phe
1 5 10 15

Gln Arg Leu Thr Pro Ser Tyr Asn Ala Asp Ile Leu Ser His Leu
20 25 30

His Tyr Leu Arg Glu Thr Pro Leu
35 40

<G1> SEQ ID NO 109
<211> LENGTH: 28
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 109

Gln Asn Ile Asp Gly Lys Leu Thr Pro Ser Tyr Asn Ala Asp Ile Leu
1 5 10 15

Leu Ser His Leu His Tyr Leu Arg Glu Thr Pro Leu
20 25

<G1> SEQ ID NO 110
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 110

Glu Thr Pro Leu
1

<G1> SEQ ID NO 111
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 111

Gln Asn Ile Asp Gly Lys Leu Thr Pro Leu
1 5 10

<G1> SEQ ID NO 112
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 112

Glu Thr Pro Leu
Glu Thr Pro Leu
1

Gln Asn Ile Asp Gly Lys Leu Thr Pro Ser Tyr Asn Ala Asp Ile Leu
1  5  10  15
Leu Ser His Leu His Tyr Leu Arg Glu Thr Pro Leu
20  25

Glu Thr Pro Leu
1

Gln Asn Ile Asp Gly Lys Leu Thr Pro Ser Tyr Asn Ala Asp Ile Leu
1  5  10  15
Leu Ser His Leu His Tyr Leu Arg Glu Thr Pro Leu
20  25

Glu Thr Pro Leu
1
<table>
<thead>
<tr>
<th>SEQ ID NO</th>
<th>LENGTH</th>
<th>TYPE</th>
<th>ORGANISM</th>
</tr>
</thead>
<tbody>
<tr>
<td>119</td>
<td>6</td>
<td>PRT</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>120</td>
<td>20</td>
<td>PRT</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>121</td>
<td>40</td>
<td>PRT</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>122</td>
<td>9</td>
<td>PRT</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>123</td>
<td>10</td>
<td>PRT</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>124</td>
<td>17</td>
<td>PRT</td>
<td>Homo sapiens</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gln Asn Ile Asp Gly Lys</td>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>Pro Ser Tyr Asn Ala Asp Ile Leu Leu Ser His Leu His Tyr Leu Arg</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu Thr Pro Leu</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro His Leu Thr Ser Asp Asp Val Asp Lys Ala Leu Gln Asn Ser Pro</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg Leu Met His Ala Arg Thr Gly Gly Ala Ala Phe Ile Phe Pro</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn Thr Ser Val Tyr Pro Glu Ala</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro Arg Leu Met His Ala Arg Asn Thr</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro His Leu Thr Ser Asp Asp Val Asp Lys</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro His Leu Thr Ser Asp Asp Val Asp Lys Ala Leu Gln Asn Ser Pro</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Pro His Leu Thr Ser Asp Asp Val Asp Lys Ala Leu Gln Asn Ser Pro
1 5 10 15

Arg

<210> SEQ ID NO 132
<211> LENGTH: 3
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 132

Arg Asn Thr
1

<210> SEQ ID NO 133
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 133

Pro His Leu Thr Ser Asp Asp Val Asp Lys Ala Leu Gln Asn Ser Pro
1 5 10 15

Arg

<210> SEQ ID NO 134
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 134

Pro His Leu Thr Ser Asp Asp Val Asp Lys Ala Leu Gln Asn Ser Pro
1 5 10 15

Arg Leu

<210> SEQ ID NO 135
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 135

Thr Gln Arg Ile Thr Thr Arg Pro Asp Leu Pro Tyr Glu Pro Pro Arg
1 5 10 15

Arg Ser Ala Trp Thr Gly His Gly His Pro Thr Pro Gln Ser Lys Ala
20 25 30

Ala Gln Pro Ser Pro Ser Thr Val
35 40

<210> SEQ ID NO 136
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 136

Asp Leu Pro Tyr Glu Pro Pro Arg
1 5

<210> SEQ ID NO 137
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 137

Ser Ala Trp Thr Gly His Gly His Pro Thr Pro Gln Ser Lys Ala Ala
Gln Pro Ser Pro Ser Thr Val

<210> SEQ ID NO 138
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400>SEQUENCE: 139
Pro Tyr Glu Pro Pro Arg Arg
1  5

<210> SEQ ID NO 139
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400>SEQUENCE: 139
Ser Ala Trp Thr Gly His Gly His Pro Thr Pro Gln Ser Lys Ala Ala
1  5  10  15
Gln Pro Ser Pro Ser Thr Val

<210> SEQ ID NO 140
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400>SEQUENCE: 140
Ser Ala Trp Thr Gly His Gly His Pro Thr Pro Gln Ser Lys Ala Ala
1  5  10  15
Gln Pro Ser Pro Ser Thr Val

<210> SEQ ID NO 141
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400>SEQUENCE: 141
Pro Tyr Glu Pro Pro Arg Arg Ser Ala Trp Thr Gly His Gly His Pro
1  5  10  15
Thr Pro Gln Ser Lys Ala Ala Gln Pro Ser Pro Ser Thr Val
20  25  30

<210> SEQ ID NO 142
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400>SEQUENCE: 142
Ser Ala Trp Thr Gly His Gly His Pro Thr Pro Gln Ser Lys Ala Ala
1  5  10  15
Gln Pro Ser Pro Ser Thr Val

<210> SEQ ID NO 143
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400>SEQUENCE: 143
Ser Ala Trp Thr Gly His Gly His Pro Thr Pro Gln Ser Lys Ala Ala
1 5 10 15

Gln Pro Ser Pro Ser Thr Val
20

<210> SEQ ID NO 144
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 144
Asp Leu Pro Tyr Glu Pro Pro Arg Arg
1 5

<210> SEQ ID NO 145
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 145
Pro Tyr Glu Pro Pro Arg Arg Ser Ala Trp Thr Gly His Gly His Pro
1 5 10 15
Thr Pro Gln Ser Lys Ala Ala Gln Pro Ser Pro Ser Thr Val
20 25 30

<210> SEQ ID NO 146
<211> LENGTH: 32
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 146
Asp Leu Pro Tyr Glu Pro Pro Arg Arg Ser Ala Trp Thr Gly His Gly
1 5 10 15
His Pro Thr Pro Gln Ser Lys Ala Ala Gln Pro Ser Pro Ser Thr Val
20 25 30

<210> SEQ ID NO 147
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 147
Asp Leu Pro Tyr Glu Pro Pro Arg Arg
1 5

<210> SEQ ID NO 148
<211> LENGTH: 32
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 148
Asp Leu Pro Tyr Glu Pro Pro Arg Arg Ser Ala Trp Thr Gly His Gly
1 5 10 15
His Pro Thr Pro Gln Ser Lys Ala Ala Gln Pro Ser Pro Ser Thr Val
20 25 30

<210> SEQ ID NO 149
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 149
 Ala Ala Gln Pro Ser Pro Ser Thr Val
1 5
<210> SEQ ID NO 150
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 150

Ser Ala Trp Thr Gly His Gly His Pro Thr Pro Gln Ser Lys Ala
1    5    10    15
Gln Pro Ser Pro Ser Thr Val
20

<210> SEQ ID NO 151
<211> LENGTH: 32
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 151

Asp Leu Pro Tyr Glu Pro Pro Arg Arg Ser Ala Trp Thr Gly His Gly
1    5    10    15
His Pro Thr Pro Gln Ser Lys Ala Ala Gln Pro Ser Pro Ser Thr Val
20    25    30

<210> SEQ ID NO 152
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 152

Pro Lys Thr Glu Asp Gln Arg Pro Gln Leu Asp Pro Tyr Gln Ile Leu
1    5    10    15
Gly Pro Thr Ser Ser Arg Leu Ala Asn Pro Gly Ser Gly Gln Ile Gln
20    25    30
Leu Thr Gln Phe Leu Leu Glu Leu
35    40

<210> SEQ ID NO 153
<211> LENGTH: 2
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 153

Pro Lys
1

<210> SEQ ID NO 154
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 154

Thr Glu Asp Gln Arg Pro Gln Leu Asp Pro Tyr Gln Ile Leu Gly Pro
1    5    10    15
Thr Ser Ser Arg
20

<210> SEQ ID NO 155
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 155

Pro Lys Thr Glu Asp Gln Arg Pro Gln Leu Asp Pro Tyr Gln Ile Leu
Gly Pro Thr Ser Ser Arg

20

<210> SEQ ID NO 156
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 156

Pro Lys Thr Glu Asp Gln Arg Pro Gln Leu Asp Pro Tyr Gln Ile Leu
1  5  10  15

Gly Pro Thr Ser Ser Arg

20

<210> SEQ ID NO 157
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 157

Pro Lys Thr Glu Asp Gln Arg Pro Gln Leu Asp Pro Tyr Gln Ile Leu
1  5  10  15

Gly Pro Thr Ser Ser Arg

20

<210> SEQ ID NO 158
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 158

Pro Lys Thr Glu Asp Gln Arg Pro Gln Leu Asp Pro Tyr Gln Ile Leu
1  5  10  15

Gly Pro Thr Ser Ser Arg

20

<210> SEQ ID NO 159
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 159

Pro Lys Thr Glu Asp Gln Arg Pro Gln Leu Asp Pro Tyr Gln Ile Leu
1  5  10  15

Gly Pro Thr Ser Ser Arg

20

<210> SEQ ID NO 160
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 160

Thr Glu Asp Gln Arg Pro Gln Leu Asp Pro Tyr Gln Ile Leu Gly Pro
1  5  10  15

Thr Ser Ser Arg

20

<210> SEQ ID NO 161
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
Pro Lys Thr Glu Asp Gln Arg Pro Gln Leu Asp Pro Tyr Gln Ile Leu
1 5 10 15
Gly Pro Thr Ser Ser Arg
20

<210> SEQ ID NO 162
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 162
Pro Lys Thr Glu Asp Gln Arg Pro Gln Leu Asp Pro Tyr Gln Ile Leu
1 5 10 15
Gly Pro Thr Ser Ser Arg
20

<210> SEQ ID NO 163
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 163
Pro Lys Thr Glu Asp Gln Arg Pro Gln Leu Asp Pro Tyr Gln Ile Leu
1 5 10 15
Gly Pro Thr Ser Ser Arg
20

<210> SEQ ID NO 164
<211> LENGTH: 2
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 164
Pro Lys
1

<210> SEQ ID NO 165
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 165
Pro Lys Thr Glu Asp Gln Arg Pro Gln Leu Asp Pro Tyr Gln Ile Leu
1 5 10 15
Gly Pro Thr Ser Ser Arg
20

<210> SEQ ID NO 166
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 166
Pro Lys Thr Glu Asp Gln Arg Pro Gln Leu Asp Pro Tyr Gln Ile Leu
1 5 10 15
Gly Pro Thr Ser Ser Arg
20

<210> SEQ ID NO 167
<211> LENGTH: 2
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 167

Pro Lys
1

<210> SEQ ID NO 168
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 168

Leu Ser Asp Ser Ser Asn Ser Ser Cys Ile Thr Trp Glu Gly Thr Asn
1  5  10  15

Gly Glu Phe Lys Met Thr Asp Pro Asp Glu Val Ala Arg Arg Trp Gly
20 25 30

Glu Arg Lys Ser Lys Pro Asn Met
35 40

<210> SEQ ID NO 169
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 169

Met Thr Asp Pro Asp Glu Val Ala Arg
1  5

<210> SEQ ID NO 170
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 170

Met Thr Asp Pro Asp Glu Val Ala Arg
1  5

<210> SEQ ID NO 171
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 171

Met Thr Asp Pro Asp Glu Val Ala Arg
1  5

<210> SEQ ID NO 172
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 172

Met Thr Asp Pro Asp Glu Val Ala Arg
1  5

<210> SEQ ID NO 173
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 173

Met Thr Asp Pro Asp Glu Val Ala Arg
1  5
<210> SEQ ID NO 174
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 174

Met Thr Asp Pro Asp Glu Val Ala Arg Arg
1 5      10

<210> SEQ ID NO 175
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 175

Thr Asp Pro Asp Glu Val Ala Arg Arg Lys Ser Lys Pro Asn Met
1 5 10      15

<210> SEQ ID NO 176
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 176

Met Thr Asp Pro Asp Glu Val Ala Arg
1 5

<210> SEQ ID NO 177
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 177

Lys Ser Lys Pro Asn Met
1 5

<210> SEQ ID NO 178
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 178

Asn Tyr Asp Lys Leu Ser Arg Ala Leu Arg Tyr Tyr Asp Lys Asn
1 5 10      15
Ile Met Thr Lys Val His Gly Lys Arg Tyr Ala Tyr Lys Phe Asp Phe
20 25     30
His Gly Ile Ala Gln Ala Leu Gln
35 40

<210> SEQ ID NO 179
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 179

Phe Asp Phe His Gly Ile Ala Gln Ala Leu Gln
1 5      10

<210> SEQ ID NO 180
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 180

Phe Asp Phe His Gly Ile Ala Gln Ala Leu Gln
<210> SEQ ID NO 181
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 181
Phe Asp Phe His Gly Ile Ala Gln Ala Leu Gln
1 5 10

<210> SEQ ID NO 182
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 182
Phe Asp Phe His Gly Ile Ala Gln Ala Leu Gln
1 5 10

<210> SEQ ID NO 183
<211> LENGTH: 25
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 183
Tyr Tyr Tyr Asp Lys Asn Ile Met Thr Lys Tyr Ala Tyr Lys Phe Asp
1 5 10 15
Phe His Gly Ile Ala Gln Ala Leu Gln
20 25

<210> SEQ ID NO 184
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 184
Asn Tyr Asp Lys Leu Ser Arg
1 5

<210> SEQ ID NO 185
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 185
Asn Tyr Asp Lys Leu Ser Arg Tyr Tyr Asp Lys Asn Ile Met Thr
1 5 10 15
Lys

<210> SEQ ID NO 186
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 186
Pro His Pro Pro Glu Ser Ser Leu Tyr Lys Tyr Tyr Pro Ser Asp Leu Pro
1 5 10 15
Tyr Met Gly Ser Tyr His Ala His Pro Gin Lys Met Asn Phe Val Ala
20 25 30
Pro His Pro Pro Ala Leu Pro Val
35 40
<210> SEQ ID NO 187
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 187
Pro His Pro Pro Glu Ser Ser Leu Tyr Lys
1  5  10

<210> SEQ ID NO 188
<211> LENGTH: 24
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 188
Pro His Pro Pro Glu Ser Ser Leu Tyr Lys Tyr Pro Ser Asp Leu Pro
1  5  10  15
Tyr Met Gly Ser Tyr His Ala His
20

<210> SEQ ID NO 189
<211> LENGTH: 27
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 189
Pro His Pro Pro Glu Ser Ser Leu Tyr Lys Tyr Pro Ser Asp Leu Pro
1  5  10  15
Tyr Met Gly Ser Tyr His Ala His Pro Gln Lys
20  25

<210> SEQ ID NO 190
<211> LENGTH: 27
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 190
Pro His Pro Pro Glu Ser Ser Leu Tyr Lys Tyr Pro Ser Asp Leu Pro
1  5  10  15
Tyr Met Gly Ser Tyr His Ala His Pro Gln Lys
20  25

<210> SEQ ID NO 191
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 191
Tyr Pro Ser Asp Leu Pro Tyr Met Gly Ser Tyr His Ala His Pro Gln Lys
1  5  10  15

<210> SEQ ID NO 192
<211> LENGTH: 27
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 192
Pro His Pro Pro Glu Ser Ser Leu Tyr Lys Tyr Pro Ser Asp Leu Pro
1  5  10  15
Tyr Met Gly Ser Tyr His Ala His Pro Gln Lys
20  25
Thr Ser Ser Ser Phe Phe Ala Ala Pro Arg Pro Tyr Trp Asn Ser Pro
1   5   10  15
Thr Gly Gly Ile Tyr Pro Asn Thr Arg Leu Pro Thr Ser His Met Pro
20  25  30
Ser His Leu Gly Thr Tyr Tyr 35

Asn Ser Pro Thr Gly
1   5

Ser Pro Thr Gly Gly Ile Tyr Pro Asn Thr Arg
1   5   10

Gly Gly Ala Ala Phe Ile Phe Pro Asn Thr Ser Val Tyr Pro Glu Ala
1   5   10  15
Thr Gln Arg Ile Thr Thr Thr Arg Pro
20

Val Pro Gln Gln Asp Trp Leu Ser Gln Pro
1   5   10

Val Pro Gln Gln Asp Trp Leu Ser Gln Pro
1   5   10
Met Ile Gln Thr Val Pro Asp Pro Ala Ala His Ile
1 5 10

<210> SEQ ID NO 200
<211> LENGTH: 55
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 200

Met Ala Ser Thr Ile Lys Glu Ala Leu Ser Val Val Ser Glu Asp Gln
1 5 10 15
Ser Leu Phe Glu Cys Ala Tyr Gly Thr Pro His Leu Ala Lys Thr Glu
20 25 30
Met Thr Ala Tyr Gly Gln Thr Ser Lys Met Ser Pro Arg Val Pro Gln
35 40 45
Gln Asp Trp Leu Ser Gln Pro
50 55

<210> SEQ ID NO 201
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 201

Met Ile Gln Thr Val Pro Asp Pro Ala Ala His Ile
1 5 10

<210> SEQ ID NO 202
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 202

gctgcagact tggccaaatg gac

<210> SEQ ID NO 203
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 203
tcaccaccga cggagcgtcc tta

<210> SEQ ID NO 204
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 204
accacatgaa tggatccagc gagtcct

<210> SEQ ID NO 205
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 205
accagcttgct gcatctttgc aacg

<210> SEQ ID NO 206
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 206
ctcgccgca cccctctcta
  20
<210> SEQ ID NO 207
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 207
gccagcagt gacctttgcc tgaag
  24
<210> SEQ ID NO 208
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 208
ctctctcca acatgaccacc aac
  23
<210> SEQ ID NO 209
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 209
gtctcgccgg aagatgactc tc
  22
<210> SEQ ID NO 210
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 210
catgtgaggc aatagttgga gtg
  23
<210> SEQ ID NO 211
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 211
ccatgttctg gaaagaggtgtgtcg
  26
<210> SEQ ID NO 212
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 212
ccttgagg gcaacacag a t
  21
<210> SEQ ID NO 213
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 213
gtgccgccg cagatgtgat ac
  22
<210> SEQ ID NO 214
<211> LENGTH: 21
<212> TYPE: DNA

<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 214

cgcaacgcc agctgtatcc c
  21

<210> SEQ ID NO 215
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 215
agcgtctggc cacctcactc
  19

<210> SEQ ID NO 216
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 216
atgtttttat gaccaacga gttcttgct
  30

<210> SEQ ID NO 217
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 217
atgacgggtt aagttcatga ttctgtg
  27

<210> SEQ ID NO 218
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 218
cggagcgc gtc acca ctcctc
  21

<210> SEQ ID NO 219
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 219
taactgagga cgcgtgtcct c
  22

<210> SEQ ID NO 220
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 220
cacagtgcc caaaa
  15

<210> SEQ ID NO 221
<400> SEQUENCE: 221
000

<210> SEQ ID NO 222
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 222
cggcaaaaga tattgcttaca aatttt

<210> SEQ ID NO 223
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 223

gacgactcgg tcggatgtg

<210> SEQ ID NO 224
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 224

cacgggcattt cccg

<210> SEQ ID NO 225
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 225

cggcgcagga agcctta

<210> SEQ ID NO 226
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 226

tccgtgacca cactcaaca ac

<210> SEQ ID NO 227
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 227

cagttgtgag tgaggacc

<210> SEQ ID NO 228
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 228

gctcgctccg atactattat gagaa

<210> SEQ ID NO 229
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 229

cacacaaaa cttgtagaag tascg

<210> SEQ ID NO 230
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 230
accagcc acc tctgc 16

<210> SEQ ID NO 231
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 231

ggtcgaggct tgaatattas actga 25

<210> SEQ ID NO 232
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 232
gctgccgcgtg ttttctgactgc 22

<210> SEQ ID NO 233
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 233
tcgagcc acc tcctc 15

<210> SEQ ID NO 234
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 234
Met Ile Gln Thr Val Pro Asp Pro Ala Ala His Ile
1 5 10

<210> SEQ ID NO 235
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 235
Ser Ala Trp Thr Gly His Gly His Pro Thr Pro Gln Ser Lys Ala Ala
1 5 10 15
Gln Pro Ser Pro Ser Thr Val
20

<210> SEQ ID NO 236
<211> LENGTH: 42
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<220> FEATURE:
<221> NAME/KEY: NON_CONS
<222> LOCATION: (18)...(19)

<400> SEQUENCE: 236
Met Ile Gln Thr Val Pro Asp Pro Ala Ala Ser His Ile Lys Glu Ala
1 5 10 15
Leu Ser Gly Ala Ala Phe Ile Phe Pro Asn Thr Ser Val Tyr Pro
20 25 30
Glu Ala Thr Gln Arg Ile Thr Thr Arg Pro
35 40

<210> SEQ ID NO 237
Met Ile Gln Thr Val Pro Asp Pro Ala Ser His Ile Lys Glu Ala  
1 5 10 15  

Leu Ser

<210> SEQ ID NO 238  
<211> LENGTH: 42  
<212> TYPE: PRT  
<213> ORGANISM: Homo sapiens  
<220> FEATURE:  
<221> NAME/KEY: NON_CONS  
<222> LOCATION: (10)..<(11)  
<220> FEATURE:  
<221> NAME/KEY: NON_CONS  
<222> LOCATION: (14)..<(15)  
<220> FEATURE:  
<221> NAME/KEY: NON_CONS  
<222> LOCATION: (18)..<(19)  

<400> SEQUENCE: 239  

Met Ala Ser Thr Ile Lys Glu Ala Leu Ser Met Thr Ala Ser Met Ser  
1 5 10 15  

Pro Arg Gly Gly Ala Ala Phe Ile Phe Pro Ase Thr Ser Val Tyr Pro  
20 25 30  

Glu Ala Thr Gln Arg Ile Thr Thr Arg Pro  
35 40

<210> SEQ ID NO 239  
<211> LENGTH: 18  
<212> TYPE: PRT  
<213> ORGANISM: Homo sapiens  
<220> FEATURE:  
<221> NAME/KEY: NON_CONS  
<222> LOCATION: (10)..<(11)  
<220> FEATURE:  
<221> NAME/KEY: NON_CONS  
<222> LOCATION: (14)..<(15)  

<400> SEQUENCE: 239  

Met Ala Ser Thr Ile Lys Glu Ala Leu Ser Met Thr Ala Ser Met Ser  
1 5 10 15  

Pro Arg

<210> SEQ ID NO 240  
<211> LENGTH: 479  
<212> TYPE: PRT  
<213> ORGANISM: Homo sapiens  

<400> SEQUENCE: 240  

Met Ala Ser Thr Ile Lys Glu Ala Leu Ser Val Val Ser Glu Asp Gln  
1 5 10 15  

Ser Leu Phe Glu Cys Ala Tyr Gly Thr Pro His Leu Ala Lys Thr Glu  
20 25 30  

Met Thr Ala Ser Ser Ser Ser Asp Tyr Gly Gin Thr Ser Lys Met Ser  
35 40 45  

Pro Arg Val Pro Gin Gln Asp Trp Leu Ser Gin Pro Pro Ala Arg Val  
50 55 60  

Thr Ile Lys Met Glu Cys Asn Pro Ser Gin Val Asn Gly Ser Arg Asn  
65 70 75 80  

Ser Pro Asp Glu Cys Ser Val Ala Lys Gly Gly Lys Met Val Gly Ser
Pro Asp Thr Val Gly Met Asn Tyr Gly Ser Tyr Met Glu Glu Lys His
100 105 110

Met Pro Pro Pro Asn Met Thr Thr Asn Glu Arg Arg Val Ile Val Pro
115 120 125

Ala Asp Pro Thr Leu Trp Ser Thr Asp His Val Arg Gln Trp Leu Glu
130 135 140

Trp Ala Val Lys Glu Tyr Gly Leu Pro Asp Val Asn Ile Leu Leu Phe
145 150 155 160

Gln Asn Ile Asp Gly Lys Glu Leu Cys Lys Met Thr Lys Asp Asp Phe
165 170 175

Gln Arg Leu Thr Pro Ser Tyr Asn Ala Asp Ile Leu Leu Ser His Leu
180 185 190

His Tyr Leu Arg Glu Thr Pro Leu Pro His Leu Thr Ser Asp Asp Val
195 200 205

Asp Lys Ala Leu Glu Asn Ser Pro Arg Leu Met His Ala Arg Asn Thr
210 215 220

Gly Gly Ala Ala Phe Ile Phe Pro Asn Thr Ser Val Tyr Pro Glu Ala
225 230 235 240

Thr Gln Arg Ile Thr Thr Arg Pro Asp Leu Pro Tyr Glu Pro Pro Arg
245 250 255

Arg Ser Ala Trp Thr Gly His Gly His Pro Thr Pro Gln Ser Lys Ala
260 265 270

Ala Gln Pro Ser Pro Ser Thr Val Pro Lys Thr Glu Asp Gln Arg Pro
275 280 285

Gln Leu Asp Pro Tyr Gln Ile Leu Gly Pro Thr Ser Ser Arg Leu Ala
290 295 300

Asn Pro Gly Ser Gly Gln Ile Gln Leu Trp Gln Phe Leu Leu Glu Leu
305 310 315 320

Leu Ser Asp Ser Ser Asn Ser Ser Cys Ile Thr Trp Gln Gly Thr Asn
325 330 335

Gly Glu Phe Lys Met Thr Asp Pro Asp Glu Val Ala Arg Arg Trp Gly
340 345 350

Glu Arg Lys Ser Lys Pro Asn Met Asn Tyr Asp Lys Leu Ser Arg Ala
355 360 365

Leu Arg Tyr Tyr Tyr Asp Lys Asn Ile Met Thr Lys Val His Gly Lys
370 375 380

Arg Tyr Ala Tyr Lys Phe Asp Phe His Gly Ile Ala Glu Ala Leu Gln
385 390 395 400

Pro His Pro Pro Glu Ser Ser Leu Tyr Lys Tyr Pro Ser Asp Leu Pro
405 410 415

Tyr Met Gly Ser Tyr His Ala His Pro Gin Lys Met Asn Phe Val Ala
420 425 430

Pro His Pro Pro Ala Leu Pro Val Thr Ser Ser Phe Phe Ala Ala
435 440 445

Pro Asn Pro Tyr Trp Asn Ser Pro Thr Gly Gly Ile Tyr Pro Asn Thr
450 455 460

Arg Leu Pro Thr Ser His Met Pro Ser His Leu Gly Thr Tyr Tyr
465 470 475

<210> SEQ ID NO 241
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<table>
<thead>
<tr>
<th>Gly</th>
<th>Ala</th>
<th>Ala</th>
<th>Phe</th>
<th>Ile</th>
<th>Phe</th>
<th>Pro</th>
<th>Asn</th>
<th>Thr</th>
<th>Ser</th>
<th>Val</th>
<th>Tyr</th>
<th>Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Met</th>
<th>Asp</th>
<th>Gly</th>
<th>Phe</th>
<th>Tyr</th>
<th>Asp</th>
<th>Gln</th>
<th>Gln</th>
<th>Val</th>
<th>Pro</th>
<th>Tyr</th>
<th>Met</th>
<th>Val</th>
<th>Thr</th>
<th>Asn</th>
<th>Ser</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td>15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gln</th>
<th>Arg</th>
<th>Gly</th>
<th>Arg</th>
<th>Asn</th>
<th>Cys</th>
<th>Asn</th>
<th>Gln</th>
<th>Lys</th>
<th>Pro</th>
<th>Thr</th>
<th>Asn</th>
<th>Val</th>
<th>Arg</th>
<th>Lys</th>
<th>Arg</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lys</th>
<th>Phe</th>
<th>Ile</th>
<th>Asn</th>
<th>Arg</th>
<th>Asp</th>
<th>Leu</th>
<th>Ala</th>
<th>His</th>
<th>Asp</th>
<th>Ser</th>
<th>Glu</th>
<th>Glu</th>
<th>Leu</th>
<th>Phe</th>
<th>Gln</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Asp</th>
<th>Leu</th>
<th>Ser</th>
<th>Glu</th>
<th>Leu</th>
<th>Gln</th>
<th>Gln</th>
<th>Glu</th>
<th>Thr</th>
<th>Trp</th>
<th>Leu</th>
<th>Ala</th>
<th>Gln</th>
<th>Ala</th>
<th>Ala</th>
<th>Gln</th>
<th>Val</th>
<th>Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>55</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Asp</th>
<th>Asp</th>
<th>Asp</th>
<th>Gln</th>
<th>Phe</th>
<th>Val</th>
<th>Pro</th>
<th>Asp</th>
<th>Tyr</th>
<th>Gln</th>
<th>Ala</th>
<th>Ala</th>
<th>Glu</th>
<th>Ser</th>
<th>Leu</th>
<th>Ala</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phe</th>
<th>His</th>
<th>Gly</th>
<th>Leu</th>
<th>Pro</th>
<th>Leu</th>
<th>Ala</th>
<th>Gln</th>
<th>Gln</th>
<th>Lys</th>
<th>Lys</th>
<th>Glu</th>
<th>Pro</th>
<th>His</th>
<th>Ser</th>
<th>Pro</th>
<th>Cys</th>
</tr>
</thead>
<tbody>
<tr>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ser</th>
<th>Glu</th>
<th>Ile</th>
<th>Ser</th>
<th>Ser</th>
<th>Ala</th>
<th>Cys</th>
<th>Ser</th>
<th>Gln</th>
<th>Glu</th>
<th>Gln</th>
<th>Glu</th>
<th>Gln</th>
<th>Pro</th>
<th>Phe</th>
<th>Lys</th>
<th>Phe</th>
<th>Ser</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>105</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tyr</th>
<th>Gly</th>
<th>Glu</th>
<th>Lys</th>
<th>Cys</th>
<th>Leu</th>
<th>Tyr</th>
<th>Asn</th>
<th>Val</th>
<th>Ser</th>
<th>Ala</th>
<th>Tyr</th>
<th>Asp</th>
<th>Gln</th>
<th>Lys</th>
<th>Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>115</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>120</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gln</th>
<th>Val</th>
<th>Gln</th>
<th>Met</th>
<th>Arg</th>
<th>Pro</th>
<th>Ser</th>
<th>Asn</th>
<th>Pro</th>
<th>Pro</th>
<th>Thr</th>
<th>Pro</th>
<th>Ser</th>
<th>Ser</th>
<th>Thr</th>
<th>Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Val</th>
<th>Ser</th>
<th>Pro</th>
<th>Leu</th>
<th>His</th>
<th>His</th>
<th>Ala</th>
<th>Ser</th>
<th>Pro</th>
<th>Asn</th>
<th>Ser</th>
<th>Thr</th>
<th>His</th>
<th>Thr</th>
<th>Pro</th>
<th>Lys</th>
</tr>
</thead>
<tbody>
<tr>
<td>145</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pro</th>
<th>Asp</th>
<th>Arg</th>
<th>Ala</th>
<th>Phe</th>
<th>Pro</th>
<th>Ala</th>
<th>His</th>
<th>Leu</th>
<th>Pro</th>
<th>Pro</th>
<th>Ser</th>
<th>Gln</th>
<th>Ser</th>
<th>Ile</th>
<th>Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>165</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Asp</th>
<th>Ser</th>
<th>Ser</th>
<th>Tyr</th>
<th>Pro</th>
<th>Met</th>
<th>Asp</th>
<th>His</th>
<th>Arg</th>
<th>Phe</th>
<th>Arg</th>
<th>Arg</th>
<th>Gln</th>
<th>Leu</th>
<th>Ser</th>
<th>Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>180</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pro</th>
<th>Cys</th>
<th>Asn</th>
<th>Ser</th>
<th>Phe</th>
<th>Pro</th>
<th>Leu</th>
<th>Pro</th>
<th>Thr</th>
<th>Met</th>
<th>Pro</th>
<th>Arg</th>
<th>Glu</th>
<th>Gly</th>
<th>Arg</th>
</tr>
</thead>
<tbody>
<tr>
<td>195</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pro</th>
<th>Met</th>
<th>Tyr</th>
<th>Gln</th>
<th>Arg</th>
<th>Gln</th>
<th>Met</th>
<th>Ser</th>
<th>Glu</th>
<th>Pro</th>
<th>Asn</th>
<th>Ile</th>
<th>Pro</th>
<th>Phe</th>
<th>Pro</th>
<th>Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>210</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gln</th>
<th>Gly</th>
<th>Phe</th>
<th>Lys</th>
<th>Gln</th>
<th>Glu</th>
<th>Tyr</th>
<th>His</th>
<th>Asp</th>
<th>Pro</th>
<th>Val</th>
<th>Tyr</th>
<th>Glu</th>
<th>His</th>
<th>Asn</th>
<th>Thr</th>
</tr>
</thead>
<tbody>
<tr>
<td>225</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Met</th>
<th>Val</th>
<th>Gln</th>
<th>Ser</th>
<th>Ala</th>
<th>Ala</th>
<th>Ser</th>
<th>Gln</th>
<th>Ser</th>
<th>Phe</th>
<th>Pro</th>
<th>Pro</th>
<th>Pro</th>
<th>Leu</th>
<th>Met</th>
<th>Ile</th>
</tr>
</thead>
<tbody>
<tr>
<td>245</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lys</th>
<th>Gln</th>
<th>Glu</th>
<th>Pro</th>
<th>Arg</th>
<th>Asp</th>
<th>Phe</th>
<th>Ala</th>
<th>Tyr</th>
<th>Asp</th>
<th>Ser</th>
<th>Glu</th>
<th>Val</th>
<th>Pro</th>
<th>Ser</th>
<th>Cys</th>
</tr>
</thead>
<tbody>
<tr>
<td>260</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>His</th>
<th>Ser</th>
<th>Ile</th>
<th>Tyr</th>
<th>Met</th>
<th>Arg</th>
<th>Gln</th>
<th>Gly</th>
<th>Phe</th>
<th>Leu</th>
<th>Ala</th>
<th>His</th>
<th>Pro</th>
<th>Ser</th>
<th>Arg</th>
</tr>
</thead>
<tbody>
<tr>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Thr</th>
<th>Glu</th>
<th>Gly</th>
<th>Cys</th>
<th>Met</th>
<th>Phe</th>
<th>Glu</th>
<th>Lys</th>
<th>Gly</th>
<th>Pro</th>
<th>Arg</th>
<th>Glu</th>
<th>Phe</th>
<th>Tyr</th>
<th>Asp</th>
<th>Asp</th>
</tr>
</thead>
<tbody>
<tr>
<td>290</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Thr</th>
<th>Cys</th>
<th>Val</th>
<th>Val</th>
<th>Pro</th>
<th>Glu</th>
<th>Lys</th>
<th>Phe</th>
<th>Asp</th>
<th>Gly</th>
<th>Asp</th>
<th>Ile</th>
<th>Lys</th>
<th>Gln</th>
<th>Glu</th>
<th>Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gly</th>
<th>Met</th>
<th>Tyr</th>
<th>Arg</th>
<th>Glu</th>
<th>Gly</th>
<th>Pro</th>
<th>Thr</th>
<th>Tyr</th>
<th>Glu</th>
<th>Arg</th>
<th>Arg</th>
<th>Gly</th>
<th>Ser</th>
<th>Leu</th>
<th>Gln</th>
</tr>
</thead>
<tbody>
<tr>
<td>325</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Leu</th>
<th>Trp</th>
<th>Gln</th>
<th>Phe</th>
<th>Leu</th>
<th>Val</th>
<th>Ala</th>
<th>Leu</th>
<th>Leu</th>
<th>Asp</th>
<th>Arg</th>
<th>Pro</th>
<th>Ser</th>
<th>Asn</th>
<th>Ser</th>
<th>His</th>
</tr>
</thead>
<tbody>
<tr>
<td>340</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phe</th>
<th>Ile</th>
<th>Ala</th>
<th>Trp</th>
<th>Thr</th>
<th>Gly</th>
<th>Arg</th>
<th>Met</th>
<th>Glu</th>
<th>Phe</th>
<th>Lys</th>
<th>Leu</th>
<th>Ile</th>
<th>Glu</th>
<th>Pro</th>
<th></th>
</tr>
</thead>
</table>
<210> SEQ ID NO 243
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 243

Pro His Ser Pro Cys Ser Glu Ile Ser Ser Ala Cys Ser Gln
1  5  10

<210> SEQ ID NO 244
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 244

Thr Pro Ser Ser Thr Pro Val Ser Pro Leu His His Ala
1  5  10

<210> SEQ ID NO 245
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 245

Arg Ala Phe Pro Ala His Leu Pro Pro Ser Glu Ser Ile Pro Asp Ser
1  5  10  15

<210> SEQ ID NO 246
<211> LENGTH: 452
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 246

Met Asp Gly Thr Ile Lys Glu Ala Leu Ser Val Val Ser Asp Asp Gln
1  5  10  15

Ser Leu Phe Asp Ser Ala Tyr Gly Ala Ala Ala His Leu Pro Lys Ala
20  25  30

Asp Met Thr Ala Ser Gly Ser Pro Tyr Gly Gln Pro His Lys Ile
35  40  45

Asn Pro Leu Pro Pro Gln Glu Trp Ile Asn Gln Pro Val Arg Val
50  55  60

Asn Val Lys Arg Glu Tyr Asp His Met Asn Gly Ser Arg Glu Ser Pro
65  70  75  80

Val Asp Cys Ser Val Ser Lys Cys Ser Lys Leu Val Gly Gly Gly Glu
<table>
<thead>
<tr>
<th>85</th>
<th>90</th>
<th>95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser Asn Pro Met Asn Tyr Asn Ser Tyr Met Asp Glu Lys Asn Gly Pro</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>Pro Pro Pro Asn Met Thr Thr Asm Glu Arg Arg Val Ile Val Pro Ala</td>
<td></td>
<td>115</td>
</tr>
<tr>
<td>Asp Pro Thr Leu Trp Thr Glu His Val Arg Glu Trp Leu Glu Trp</td>
<td></td>
<td>130</td>
</tr>
<tr>
<td>Ala Ile Lys Glu Tyr Ser Leu Met Glu Ile Asp Thr Ser Phe Phe Glu</td>
<td></td>
<td>145</td>
</tr>
<tr>
<td>Asn Met Asp Gly Lys Glu Leu Cys Lys Met Aen Lys Glu Asp Phe Leu</td>
<td></td>
<td>165</td>
</tr>
<tr>
<td>Arg Ala Thr Thr Leu Tyr Asn Thr Glu Val Leu Leu Ser His Leu Ser</td>
<td></td>
<td>180</td>
</tr>
<tr>
<td>Tyr Leu Arg Glu Ser Ser Leu Leu Ala Tyr Asn Thr Thr Ser His Thr</td>
<td></td>
<td>195</td>
</tr>
<tr>
<td>Asp Gln Ser Ser Arg Leu Ser Val Lys Glu Asp Pro Ser Tyr Asp Ser</td>
<td></td>
<td>210</td>
</tr>
<tr>
<td>Val Arg Arg Gly Ala Trp Gly Asn Aen Met Aen Ser Gly Leu Aen Lys</td>
<td></td>
<td>225</td>
</tr>
<tr>
<td>Ser Pro Pro Leu Gly Gly Ala Gln Thr Ile Ser Lys Aen Thr Glu Gln</td>
<td></td>
<td>245</td>
</tr>
<tr>
<td>Arg Pro Gln Pro Asp Pro Tyr Gly Ile Leu Gly Pro Thr Ser Ser Arg</td>
<td></td>
<td>260</td>
</tr>
<tr>
<td>Leu Ala Aen Pro Gly Ser Gly Gln Ile Gln Leu Trp Glu Phe Leu Leu</td>
<td></td>
<td>275</td>
</tr>
<tr>
<td>Glu Leu Leu Ser Asp Ser Ala Aen Ala Ser Cys Ile Thr Trp Glu Gly</td>
<td></td>
<td>290</td>
</tr>
<tr>
<td>Thr Aen Gly Glu Phe Lys Met Thr Asp Pro Asp Glu Val Ala Arg Arg</td>
<td></td>
<td>305</td>
</tr>
<tr>
<td>Trp Gly Glu Arg Lys Ser Lys Pro Aen Met Aen Tyr Asp Lys Leu Ser</td>
<td></td>
<td>325</td>
</tr>
<tr>
<td>Arg Ala Leu Arg Tyr Tyr Tyr Asp Lys Aen Met Thr Lys Val His</td>
<td></td>
<td>340</td>
</tr>
<tr>
<td>Gly Lys Arg Tyr Ala Tyr Lys Phe Asp Phe His Gly Ile Ala Gln Ala</td>
<td></td>
<td>355</td>
</tr>
<tr>
<td>Leu Gln Pro His Pro Thr Glu Ser Ser Met Tyr Lys Tyr Pro Ser Asp</td>
<td></td>
<td>370</td>
</tr>
<tr>
<td>Ile Ser Tyr Met Pro Ser Tyr His Ala His Gin Gin Lys Val Aen Phe</td>
<td></td>
<td>385</td>
</tr>
<tr>
<td>Val Pro Pro His Pro Ser Met Pro Val Thr Ser Ser Ser Phe Phe</td>
<td></td>
<td>405</td>
</tr>
<tr>
<td>Gly Ala Ala Ser Gin Tyr Trp Thr Ser Pro Thr Gly Gly Ile Tyr Pro</td>
<td></td>
<td>420</td>
</tr>
<tr>
<td>Aen Pro Aen Val Pro Arg His Pro Aen Thr His Val Pro Ser His Leu</td>
<td></td>
<td>435</td>
</tr>
<tr>
<td>Gly Ser Tyr Tyr</td>
<td></td>
<td>450</td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 247
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 247

Ser Leu Phe Asp Ser Ala Tyr Gly Ala Ala Ala His Leu Pro
His Lys Ile Asn Pro Leu Pro Pro Gln Ile Asn Gln Pro Val Arg Val
1 5 10 15

Asn Val

Arg Glu Ser Pro Val Asp Cys Ser Val Ser Lys Cys Ser Lys Leu Val
1 5 10 15

Gly

Thr Gln Glu His Val Arg Gln
1 5

Ser Ser Arg Leu Ser Val Lys Glu
1 5

Pro Asn Thr His Val Pro Ser His Leu
1 5

Met Glu Arg Arg Met Lys Ala Gly Tyr Leu Asp Gln Gln Val Pro Tyr
1 5 10 15

Thr Phe Ser Ser Lys Ser Pro Gly Asn Gly Ser Leu Arg Glu Ala Leu
20 25 30

Ile Gly Pro Leu Gly Lys Leu Met Asp Pro Gly Ser Leu Pro Pro Leu
35 40 45

Asp Ser Glu Asp Leu Phe Gln Asp Leu Ser His Phe Gln Glu Thr Trp
50 55 60
Leu Ala Glu Ala Gln Val Pro Asp Ser Asp Glu Gln Phe Val Pro Asp

65  70  75  80
Phe His Ser Glu Asn Leu Ala Phe His Ser Pro Thr Thr Arg Ile Lys

95  90  95
Lys Glu Pro Gin Ser Pro Pro Thr Asp Pro Ala Leu Ser Cys Ser Arg

100 105 110
Lys Pro Pro Leu Pro Tyr His His Gly Glu Gin Cys Leu Tyr Ser Ser

115 120 125
Ala Tyr Asp Pro Pro Arg Gin Ile Ala Ile Lys Ser Pro Ala Pro Gly

130 135 140
Ala Leu Gly Gin Ser Pro Leu Gin Pro Phe Pro Arg Ala Glu Gin Arg

145 150 155 160
Asn Phe Leu Arg Ser Ser Gly Thr Ser Gin Pro His Pro Gly His Gly

165 170 175
Tyr Leu Gly Glu His Ser Ser Val Phe Gin Gin Pro Leu Asp Ile Cys

180 185 190
His Ser Phe Thr Ser Gin Gly Gly Gly Arg Glu Pro Leu Pro Ala Pro

195 200 205
Tyr Gin His Gin Leu Ser Gin Pro Cys Pro Pro Tyr Pro Gin Gin Ser

210 215 220
Phe Lys Gin Glu Tyr His Asp Pro Leu Tyr Glu Gin Ala Gly Gin Pro

225 230 235 240
Ala Val Asp Gin Gly Gly Val Asn Gly His Arg Tyr Pro Gly Ala Gly

245 250 255
Val Val Ile Lys Gin Glu Thr Asp Phe Ala Tyr Asp Ser Asp Val

260 265 270
Thr Gly Cys Ala Ser Met Tyr Leu His Thr Glu Gly Phe Ser Gly Pro

275 280 285
Ser Pro Gly Asp Gly Ala Met Gly Tyr Gly Tyr Glu Lys Pro Leu Arg

290 295 300
Pro Phe Pro Asp Val Cys Val Val Pro Glu Lys Phe Glu Gly Asp

305 310 315 320
Ile Lys Gin Glu Gly Val Gly Ala Phe Arg Glu Gly Pro Pro Tyr Gin

325 330 335
Arg Arg Gin Ala Leu Gin Leu Trp Gin Phe Leu Val Ala Leu Leu Gin

340 345 350
Asp Pro Thr Asn Ala His Phe Ile Ala Trp Thr Gly Arg Gly Met Glu

355 360 365
Phe Lys Leu Ile Glu Pro Glu Val Ala Arg Leu Trp Gly Ile Gin

370 375 380
Lys Asn Gin Pro Ala Gin Ala Gin Gin Gin Gin Gin Gin Gin Gin Gin

395 395 400
Tyr Tyr Tyr Glu Lys Gly Lys Val Ala Gly Glu Arg Tyr

405 410 415
Val Tyr Lys Phe Val Cys Glu Pro Glu Ala Leu Phe Ser Leu Ala Phe

420 425 430
Pro Asp Asn Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin

440 445
Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin

450 455 460
Tyr Leu Pro Glu Leu Ala Gly Pro Ala Gin Gin Gin Gin Gin Gin Gin

465 470 475 480
Gly Tyr Ser Tyr
<table>
<thead>
<tr>
<th>SEQ ID NO</th>
<th>LENGTH</th>
<th>TYPE</th>
<th>ORGANISM</th>
</tr>
</thead>
<tbody>
<tr>
<td>254</td>
<td>12</td>
<td>PRT</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>255</td>
<td>11</td>
<td>PRT</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>256</td>
<td>7</td>
<td>PRT</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>257</td>
<td>7</td>
<td>PRT</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>258</td>
<td>20</td>
<td>PRT</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>259</td>
<td>25</td>
<td>PRT</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>260</td>
<td>5</td>
<td>PRT</td>
<td>Homo sapiens</td>
</tr>
</tbody>
</table>

**SEQUENCE: 254**

Gly Tyr Leu Amp Gln Gln Val Pro Tyr Thr Phe Ser

| 1 | 5 | 10 |

**SEQUENCE: 255**

Leu Arg Glu Ala Leu Ile Gly Pro Leu Gly Lys

| 1 | 5 | 10 |

**SEQUENCE: 256**

Leu Phe Glu Asp Leu Ser His

| 1 | 5 |

**SEQUENCE: 257**

Ser Glu Asn Leu Ala Phe His

| 1 | 5 |

**SEQUENCE: 258**

Thr Asp Pro Ala Leu Ser Cys

| 1 | 5 |

**SEQUENCE: 259**

Ser Arg Lys Pro Pro Leu Pro Tyr His His Gly Glu Gln Cys Leu Tyr

| 1 | 5 | 10 | 15 |

**SEQUENCE: 260**

Asp Pro Pro Arg Glu Ile Ala Ile Lys Ser Pro Ala Pro Gly Ala Leu

| 1 | 5 | 10 | 15 |
Gly Glu Ser Pro Leu Glu Pro Phe Pro
20 25

<210> SEQ ID NO 261
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 261
Ser Glu Pro His Pro Gly His Tyr Leu Gly Glu His Ser Ser Val Phe
1 5 10 15

Gln Gln Pro Leu Asp
20

<210> SEQ ID NO 262
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 262
Ile Cys His Ser Phe
1 5

<210> SEQ ID NO 263
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 263
Glu Pro Leu Pro Ala Pro Tyr Gln His Gln Leu Ser Glu Pro Cys Pro
1 5 10 15

Pro Tyr Pro Gln Gln
20

<210> SEQ ID NO 264
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 264
Tyr His Asp Pro Leu Tyr Glu Gln Gly Glu Gln Pro Ala Val Asp Gln
1 5 10 15

<210> SEQ ID NO 265
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 265
Arg Tyr Pro Gly Ala Gly Val Val Ile Lys
1 5 10

<210> SEQ ID NO 266
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 266
Tyr Asp Ser Asp Val Thr Gly Cys Ala Ser Met Tyr
1 5 10

<210> SEQ ID NO 267
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 267

Tyr Glu Lys Pro Leu Arg Pro Phe Pro Asp Asp Val
1  5

<210> SEQ ID NO 268
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 268

Cys Val Val Pro Glu
i  5

<210> SEQ ID NO 269
<211> LENGTH: 71
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 269

cgcagctaa gcagagggcg gaggcgaggg cggagggcga ggggggggga ggcgctcg
60
gagcgccgca g
71

<210> SEQ ID NO 270
<211> LENGTH: 142
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 270

cgcagctaa gcagagggcg gaggcgaggg cggagggcga ggggggggga ggcgctcg
60
gagcgccgca ggtctatatg aacattccag atacctatac ttactcgtag ctgttgataa
120
cagcaagatg gttttgaaact ca
142

<210> SEQ ID NO 271
<211> LENGTH: 365
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 271

cgcagctaa gcagagggcg gaggcgaggg cggagggcga ggggggggga ggcgctcg
60
gagcgccgca ggtctatatg aacattccag atacctatac ttactcgtag ctgttgataa
120
cagcaagatg gttttgaaact cagcgctacc accagctatt ggaccttact atgaaaca
180
tggataccaa cggaaaaacc cctatccgca acagccccat gttgccccca cttctcaag
240
gttgcctcg ggtcagtaat accgtctccc cgtgcctcag taagccccga gggtctctag
300
gcacgcttc aaaccgctcg tctgcacgca gcocaaatcc ocacgccccg cagctgtcag
360
cctca
365

<210> SEQ ID NO 272
<211> LENGTH: 452
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 272

cgcagctaa gcagagggcg gaggcgaggg cggagggcga ggggggggga ggcgctcg
60
gagcgccgca ggtctatatg aacattccag atacctatac ttactcgtag ctgttgataa
120
cagcaagatg gttttgaaact cagcgctacc accagctatt ggaccttact atgaaaca
180
tggtaccaac cggaaaacec cctatecgcg acageccact gttgcccca cttctcaega  240
ggtgcccg gcctcact accgcctcgc ctgcccccag taagccccga ggtctccgtac  300
gccgcttcc aacccgctcg ctcgcaagca gccaacatct ccattcccgaa cactggtcac  360
tctaaagact aagaaacagcg tctgctacac ctggcccctg ggagacctcc ttcgggggcg  420
tgctgctgctg cgcctgctcc tctggaagtca  452

<210> SEQ ID NO: 273
<211> LENGTH: 572
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 273
cgctggcctaa gcggagggcg gaggcgcgag ggggagggaga gggggggaga gcgcggcttg  60
gagcggcggc gcgtcatacg gacatcagcc ttacacatca ttactcgtag ctttgataaa  120
cacgcaagact gctttgacat caggctaccc accagcttatt ggcaccttact atgaaacacc  180
tggctaaac cggaaacacc ccctatcgcgc acacccgacac cttgcccaca cttgctccag  240
gctggccaccg gcctgcctcct acgctgcccc cttgcccacg ggtctccgct gtcctcagc  300
gcggcccttc aacccgctcg ctcgacccga gccaacatcc ccattcggga cgcctcagc  360
tctaaagact aagaaacagcg tctgctacac ctggcccctg ggagacctcc ttcgggggcg  420
tgctgctgctg cgcctgctcc tctggaagtca cagttggcagc aagttggcaca aaccttggtat  480
agctggccgc gtcgctggctc tttggaacctttt tctggcccct cccttgctaa ctcgctcaca  540
ttcggggcgc ggggagggagc aagatcgtgct tg  572

<210> SEQ ID NO: 274
<211> LENGTH: 67
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 274
ccttgataaa taagtttctaa agagggcctc cgcctacgta aagagcttttt cttccgcttt  60
cctgcag  67

<210> SEQ ID NO: 275
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 275
atgtaaaga gtttttctcc cgccttctcg cag  33

<210> SEQ ID NO: 276
<211> LENGTH: 102
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 276
gtttatccgs gtcttttgaga gagccggagga aagcctgtgct gacccaagcc aagacaaatgg  60
actcacaagga aaaaaagtgc gcagaaaccag gggcaactaa ag  102

<210> SEQ ID NO: 277
<211> LENGTH: 66
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 277
ccgtcaggtt ctgaacagct ggtagatggg ctggtctact gaagcactg attcagactg
60
tcccggaacc agcagctcatt ctcaag
86
<210> SEQ ID NO 278
<211> LENGTH: 218
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 278

gaaagccttt cagttgtagag tgaagaccag tctgctgtggg agtgtcgcata ggaacgcca
60
cactggcata agacagagat gacgcgtcct tctccagcag actatgggca gaccccaag
120
atgagccacac ggtctctcata gcagggatgg ctgtctcaac cccccagcag acgtaaccac
180
aaaatggaag ctaccccttaaa ccaagctgaa aggctcaag agctcaag
218
<210> SEQ ID NO 279
<211> LENGTH: 152
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 279
gaactctcct gtagaagcag tgggtggcaca agggcaggaag atggttgaggca gccccagaca
60
cgttggttggc aggtaagcgca gcctacttga ggaagacgac atgccaacccc caaacatgac
120
cacgaagaag cggcagagtt atctgctcagc ag
152
<210> SEQ ID NO 280
<211> LENGTH: 204
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 280

atgctcaagct atggttagcaca gacactgtgc gcgagtgggct gggtggggcg gttgaagaaat
60
atggcctctcc agaagtctcaag atcttggttat tctacagacat cgattggggaag gacctggtgca
120
agatgaccacac ggaagacgctc cgagggctca ccccccagctc caacggcgcac atctctctctt
180
ccacacctcctcactctcaggac
204
<210> SEQ ID NO 281
<211> LENGTH: 81
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 281
tacctctctc acatttgaact tcaagatgtag tttgataaaaag cttacaaaaac tctcaggtg
60
taatgcatgc tagaaccaca g
81
<210> SEQ ID NO 282
<211> LENGTH: 72
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 282

ggggtgcagc tttatatcct ccaataactt cagttacaagc tcaagatgac caagacgaat
60
catactagcc g
72
<210> SEQ ID NO 283
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
atttacata tgagcccccag aggagatcag ccctggacagg tcaacggcac ccacggcccc 60
agctgaaag 69

cgctcacc acctccttcc aca-agtgccca aaccgtgaga cccgctcct cagttag 57

gcagtgcaac gattcttgga ccaacaagta gctgcttgct gatcactcag 48

gcaagtggcca gatcagctt cggcagtcc tctctggaac cctggccagac agtccaaact 60
ccaagtgtct cacccgggaa ggcacacaccc ggagtgccaa aagtgacagg gcacggaggg 120
tgcggccgggc ctggggagaa cggaagagcc accccaaacct gaactagctg aagctcagcc 160
gctgcctctg ttctactaat gacaaggaac tcaattaccc ggtcctaggg aagogctcag 240
cctacaggt cgacatccac gggactgccc agggccttcca gccctcaccgccc acggatccat 300
cctgtacaat gtacgcttca gacccctcggt acaatgccct ctcacgacct caccocacaga 360
agatgacaccttgctggcggcac cccctccctcc gacactcttcg ccgcttttttcttatgtg 420
cgccccaa ccaatactgg aatccaccc caagggctgtat ataccccaac actaggcctac 480
ccacagccca taggcctcct cactctgggca cttactacta a 521

gcaagtggcca gatcagctt cggcagtcc tctctggaac cctggccagac agtccaaact 60
ccaagtgtct cacccgggaa ggcacacaccc ggagtgccaa aagtgacagg gcacggaggg 120
ctcaagcttt tgggtgcccc cccctccctcc gacactcttcg ccgcttttttcttatgtg 420
cgccccaa ccaatactgg aatccaccc caagggctgtat ataccccaac actaggcctac 480
ccacagccca taggcctcct cactctgggca cttactacta a 521

gcaagtggcca gatcagctt cggcagtcc tctctggaac cctggccagac agtccaaact 60
ccaagtgtct cacccgggaa ggcacacaccc ggagtgccaa aagtgacagg gcacggaggg 120
ctcaagcttt tgggtgcccc cccctccctcc gacactcttcg ccgcttttttcttatgtg 420
cgccccaa ccaatactgg aatccaccc caagggctgtat ataccccaac actaggcctac 480
ccacagccca taggcctcct cactctgggca cttactacta a 521

tgggttttca tgggtgccc cggcagtctg gaaagaacc cccccagctca tgggttgg 60
tgggttctgc tgggtgccc cggcagtctg gaaagaacc cccccagctca tgggttgg 60
tcaatgtcag ccctgctgca gctgacctt cctgctgca gctgacctt cctgctgca 120
tcaatgtcag ccctgctgca gctgacctt cctgctgca gctgacctt cctgctgca 120

<210> ORGANISM: Homo sapiens
<400> SEQ ID NO: 290
<211> LENGTH: 248
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQ ID NO: 290

atggcgccg cagctttctg aaacctctga atctctttct cccttgccgs cgatgccaag 60
ggagaaggct ctatgtagcc aaagcagctgt gcttgaccca aaacctcttcc ccocaccaca 120
agcttttaag cagagcttac agaaccagtt gatgaacac aacaactggt tgtgcaagtc 180
ggccagcaca agtcttcocct cctcttctgat gattaacac gaaaccagag atttttcata 240
tgactcag 248

<210> SEQ ID NO: 291
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQ ID NO: 291

aagtgccttg ctgccacctc atttatatga ggcaagaggg cttcctggtct catccacagc 60
gaacagaag 69

<210> SEQ ID NO: 292
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQ ID NO: 292

gctgtatggt tgaagaaggcc ccagggcagt tttatgtaga caacgtggttt gtccacagaa 60
aattcgatg 69

<210> SEQ ID NO: 293
<211> LENGTH: 170
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQ ID NO: 293

gagacatca caacagagcca ggaatgtatgc ggaagggacc cacatacaca cgggccggt 60
cacctcagct ctggcaggttt ttgtagatct tcttgagatga cccttcaaat tctctattta 120
tttgcttgac ttggctaggc atggcaatatta aactgttaga gctctgaagag 170

<210> SEQ ID NO: 294
<211> LENGTH: 162
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQ ID NO: 294

gtggccgcagctttgaggaatctgacta tcagaaaaac aggccagctg tgaactatga taaactttagag 60
cgttcacctc gctttactta tgaagaagag attatgcaggag 102
<210> SEQ ID NO 295
<211> LENGTH: 222
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 295

ttggtctggag agagatagt ctacaagttt gtttgtagtc cagaagcct ttttctcatg  60
gcctttcgcg ataatacagcg ttcactgtcg aagagaca ctaaagctca cataaagcag  120
gagacacag tggcttttcg ttaattgat gagaatcag cgtcatacgc ggaagggggc  180
tgtgcaacc ctacacccota cagaagaggct cagttgatt aa  222

<210> SEQ ID NO 296
<211> LENGTH: 93
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 296

aatacgcccg gaaatgggag cttgctgcaaa ggcgctgatcg gcgcggctgg gaagotcatg  60
gaccggggt cctgagccgc ctcgactctg aag  93

<210> SEQ ID NO 297
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 297

atcttttcca ggatctaaag cacttccagg agacgtggtg cgtgaag  48

<210> SEQ ID NO 298
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 299

cctcagtgacc agacagtgat gagcagtttg ttcctgattt cccattcagaa aacc  54

<210> SEQ ID NO 299
<211> LENGTH: 127
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 299

tagtttccca cagcggccac cagaggtaca agaaggggc cccagtcg ccgagagacc  60
cggcgcctgc cttgacagag aagccgcac cccggtacca cccgtgagag cagtcgcttt  120
accccg  127

<210> SEQ ID NO 300
<211> LENGTH: 162
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 300

tgcttatgacccccgccag aatactgcag ccaatcccttt gcgctccttg cccatggaca  60
gtcgccctta cagcgtctttt cccgggcaag gcaacggaat ttccttgact cctctgagac  120
cctccaagcc cccgctggec atggttaccc cggggaacat ag  162

<210> SEQ ID NO 301
<211> LENGTH: 266
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 301
ctcgtcctc ccagcacccc tggacatttg ccactcctc acatotcagg gagggggcgg 60
ggaacccct ccagcccccct acaacacca ggtgctcggag cccggeccac ccatcctccca 120
gcagagcttt aagcaagaatt acatgatcct cctgtatgaa caggccccggc aggagacggt 180
ggacaggtg gggtcataag gcgactagtta cccagaggcc ggggtggtgta tcaacacagga 240
acagacggcc ttcgctatacg actcag 266

<210> SEQ ID NO 302
<211> LENGTH: 75
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 302
gttgaagggc gttgcaacaag atgtactcc acacagaggg cttcctctg ggggcccctct 60
gttgcaagggc catgg 75

<210> SEQ ID NO 303
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 303
gctatggtctac tgagaacct ctgcagccatt tcctcagatag tgtgtcgcgtt gtccttgaga 60
aatattsgag 69

<210> SEQ ID NO 304
<211> LENGTH: 173
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 304
gacacgtaac cccagaggg gttgcgggcata ccgagaggg ggccgccctac cagcgcgggg 60
gttgccacac ggtggcgaata ttcttggttg ggcttgctgga tgaacccttc aagtgccattc 120
tatctggct gacgccggcg ggaatgggaatt tcaagctctat tcaagctctgc gag 173

<210> SEQ ID NO 305
<211> LENGTH: 102
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 305
gttgccggggc tctgggggtct cccagagaac cggccacgcct tgaattaaca caaggtggac 60
cgtcgtgctcc gatactattat tgaagaaggg aacatgcagca ag 102

<210> SEQ ID NO 306
<211> LENGTH: 225
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 306
gttgtgtctt cagcgttact gcacaggttt gttgtgtgac cccgcgcccc cttctctttg 60
gccctcgggcc acaactacgg tcacagcttc aaggtgagt tgaooaccgcc ggctgatgag 120
gagacacaca tccctctgctc ccacttggtat gagacgcccc cctactcccc aagatgtgct 180
gggcccggcc aagcacttggg ccccaagggg ggctactttt actag 225
<210> SEQ ID NO 307
<211> LENGTH: 3226
<222> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 307

cgcyagctaa gcggagggcg gagcggcagg cggagggcga gggggcgggga ggcogcctg 60
gagcgccgca ggtcatatgt acaatccag atacatatac ttactctag tctttgtaaa 120
cagcacaagcg ggcttgaacct cagggcctac accagttatt gacgttaaact atgaagacca 180
tggataccaa cggagaaccc cctatccgca acagccccac tgtgtcgcaca tgttcacga 240
ggtgcacccg gtcctacact acccgtccccc cgtggccccag taccggccga gggcttcggg 300
gcaggtctcc aacccgctcg tgtgcacgca gcgcacaaat cctacgccgg aagctgtgac 360
tctcaagact aagagagcga tgtgctacac cttgaccccc ggcacccgct gccgagggcg 420
tggcgtgctgc gtcggctcaac ttcgagaaatt ctagggcagg aagctgtgac acccttggtt 480
agagttggccg agatgttgcaacc cctctctgctt tgggtttgcttg ggtggtcacc 540
ttcggccggtt gggagagccag aagatgtgcttg tgtgacaccc tccgacaccg acccagttc 600
tcagatgtgct tccatctcagta gcagcctctg gcaccctctgt gcacaaagcgc actcttgaag 660
gatcagagaggg cgccggccatt gcggccatatt ataatatttt acccttggtc 720
agggagagttg gattagcagcg gattacccag ctctattgaa cttggccaccg cccgagggcg 780
tgtagcatac attataaattt ggctcctggg cagctgctcct ttccttctgatt ccctttctcc 840
tttacagctg tattggtctgg ccggctgccg ggtggctgccg acgcagcaccg acccagttg 900
cgcggagcag gcgggctcctg cggctggctctg agcctgctgac cccctcacacac cccctcttatctt 960
eaacgtgcacg aggcacgctca ctcgggccct gcagagcagcg ccccctcctctgatgctgcg 1020
ggaaacactc cttaacaaat cattgcttgg ggcgcggttg cggggagattt tgagacacatc 1080
ttcctagttc tatggagcag gcattacagc agaaagatgt atttctcagc caaatattgag 1140
tttgccacac aagagacatt acctgctgtc atgagactgc caggacccct cgccttttaa 1200
gagacatagc aacggatgct tgtgcttcacc cccgagggcg atgtgctcag cggacagcac 1260
tttgctggtatt tccggcgggg gggccagcag gccagaggg aacatctccg aagctgtgctg 1320
gctgcttacg ccggttcatt ccggcgttcg ggcgttcgcttg ccggttctcag ccctttctcc 1380
ttttgcacaa cgcggttcag tgtgctctcc gttcgcctgctg gggvacggct gttcggttttc 1440
cccttagcag cgctgctctcc cggctgcctg gggctggctg ccctttctcc cggctgcctg 1500
taaccggcag cggcgagccg cctttccctg ccgtgcctcc ggtcggtgctgc gattctggtt 1560
cccttggcag cccgagccct ggtgcatctgc ctgtggcttc ggctgtgctgc gattctggtt 1620
gcctgctggtt cgctgtggctgc gcgtgctggtc gcgtgctggtg gcgtgctggtc gcgtgctggtc 1680
ctgagctggg gggccagggag cttgccggttt ctgctgtgttg gcgtgctggtc gcgtgctggtc 1740
ccggccttggg tcgctggctc ggtctggctc ggtctggctc ggtctggctc ggtctggctc 1800
ttttgggaggg gggcgggttg cttgccggttt ctgctgtgttg gcgtgctggtc gcgtgctggtc 1860
ccggccttggg tcgctggctc ggtctggctc ggtctggctc ggtctggctc ggtctggctc 1920
ccggccttggg tcgctggctc ggtctggctc ggtctggctc ggtctggctc ggtctggctc 1980
ccggccttggg tcgctggctc ggtctggctc ggtctggctc ggtctggctc ggtctggctc 2040
ccggccttggg tcgctggctc ggtctggctc ggtctggctc ggtctggctc ggtctggctc 2100
ccggccttggg tcgctggctc ggtctggctc ggtctggctc ggtctggctc ggtctggctc 2160
We claim:
1. A composition comprising a first oligonucleotide probe comprising a sequence that hybridizes to a 5' portion of a chimeric DNA or chimeric mRNA from a transcriptional regulatory region of an androgen regulated gene, wherein the androgen regulated gene is selected from the group consisting of TMPRSS2 and PSA, and a second oligonucleotide probe comprising a sequence that hybridizes to a 5' portion of the chimeric DNA or chimeric mRNA from an ETS family member gene, wherein the ETS family member gene is selected from the group consisting of ERG, ETVI (ER81), FL11, ETS1, ETS2, ELK1, ET6 (TEL1), ET7 (TEL2), GABpX, ELF1, ET4 (ELA; PEAK), ET5 (ERM), ERF, PEAK, ELAF, PU.1, ESE1/ESX, SAP1 (ELK4), ET3 (METS), EWS/FL11, ESE1, ESE2 (ELF5), ESE3, PDEF, NET (ELK3; SAP2), NERF (ELF2), and FEV.
2. The composition of claim 1 wherein the oligonucleotide probe comprising a sequence that hybridizes to a 5' portion of a chimeric DNA or chimeric mRNA from a transcriptional regulatory region of an androgen regulated gene, wherein the androgen regulated gene is selected from the group consisting of TMPRSS2 and PSA, and the ETS family member gene is selected from the group consisting of ERG, ETVI, and ET4.
3. The composition of claim 1 wherein the androgen regulated gene is TMPRSS2 and the ETS family member gene is an ERG gene.
4. The composition of claim 1 wherein the chimeric DNA or chimeric mRNA is obtained from a tissue, blood, urine, semen, prostatic secretion, plasma, serum, urine supernatant, urine cell pellet, or prostate cell sample.
5. The composition of claim 1 wherein the composition comprises an in situ hybridization probe for detecting a deletion or rearrangement associated with a fusion of the androgen regulated gene with the ETS family member gene.
6. A kit comprising one or more composition of claim 1.
7. The kit of claim 5 comprising one or more composition selected from the group consisting of a first labeled probe having an oligonucleotide sequence that hybridizes to a transcriptional regulatory region of a TMPRSS2 gene and a second labeled probe having an oligonucleotide sequence that hybridizes to an ERG gene;
a first labeled probe having an oligonucleotide sequence that hybridizes to a transcriptional regulatory region of a TMPRSS2 gene and a second labeled probe having an oligonucleotide sequence that hybridizes to an ETVI gene; and
a first labeled probe having an oligonucleotide sequence that hybridizes to a transcriptional regulatory region of a TMPRSS2 gene and a second labeled probe having an oligonucleotide sequence that hybridizes to an ETIV gene.
8. A composition comprising a first amplification oligonucleotide comprising a sequence that hybridizes to a 5' portion of a chimeric DNA or chimeric mRNA from a transcriptional regulatory region of an androgen regulated gene, wherein the androgen regulated gene is selected from the group consisting of TMPRSS2 and PSA, and a second amplification oligonucleotide comprising a sequence that hybridizes to a 5' portion of the chimeric DNA or chimeric mRNA from an ETS family member gene, wherein the ETS family member gene is selected from the group consisting of ERG, ETVI (ER81), FL11, ETS1, ETS2, ELK1, ET6 (TEL1), ET7 (TEL2), GABpX, ELF1, ET4 (ELA; PEAK), ET5 (ERM), ERF, PEAK, ELAF, PU.1, ESE1/ESX, SAP1 (ELK4), ET3 (METS), EWS/FL11, ESE1, ESE2 (ELF5), ESE3, PDEF, NET (ELK3; SAP2), NERF (ELF2), and FEV.
9. The composition of claim 8 wherein the androgen regulated gene comprises TMPRSS2 or PSA, and the ETS family member gene comprises ERG, ETVI, or ETIV.
10. The composition of claim 8 wherein the androgen regulated gene is TMPRSS2 and the ETS family member gene is an ERG gene.
11. The composition of claim 8, wherein the chimeric DNA or chimeric mRNA is obtained from a tissue blood, urine, semen, prostatic secretion, plasma, serum, urine supernatant, urine cell pellet, or prostate cell sample.

12. A kit comprising one or more composition of claim 8.

13. The kit of claim 12 comprising one or more composition selected from the group consisting of:

a) a first amplification oligonucleotide having a sequence that hybridizes to a transcriptional regulatory region of a TMPRSS2 gene, and a second amplification oligonucleotide having a sequence that hybridizes to an ERG gene;

b) a first amplification oligonucleotide having a sequence that hybridizes to a transcriptional regulatory region of a TMPRSS2 gene, and a second amplification oligonucleotide having a sequence that hybridizes to an ETV1 gene; and

c) a first amplification oligonucleotide having a sequence that hybridizes to a transcriptional regulatory region of a TMPRSS2 gene, and a second amplification oligonucleotide having a sequence that hybridizes to an ETV4 gene.

14. A composition comprising a hybridized oligonucleotide:target gene fusion DNA or RNA duplex, wherein the target gene fusion comprises a fusion of a TMPRSS2 or a PSA gene with an ERG, an ETV1, or an ETV4 gene, wherein the oligonucleotide is hybridized to a 5' portion of a chimeric DNA or chimeric mRNA from a transcriptional regulatory region of the TMPRSS2 or the PSA gene, or it is hybridized to a 3' portion of the chimeric DNA or chimeric mRNA from gene the ERG, the ETV1, or the ETV4 gene, and wherein the target gene fusion is obtained from a tissue, blood, urine, semen, prostatic secretion, plasma, serum, urine supernatant, urine cell pellet, or prostate cell sample.

15. The composition of claim 14 wherein the target gene fusion comprises a fusion of a TMPRSS2 gene with an ERG gene.

16. The composition of claim 14 wherein the hybridized oligonucleotide is a probe.

17. The composition of claim 14 wherein the hybridized oligonucleotide is an amplification oligonucleotide.
It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Col. 255, Lines 1-4
CLAIM 11 reads: “The composition of claim 8, wherein the chimeric DNA or chimeric mRNA is obtained from a tissue blood, urine, semen, prostatic secretion, plasma, serum, urine supernatant, urine cell pellet, or prostate cell sample.”

when in fact it should read:
“The composition of claim 8, wherein the chimeric DNA or chimeric mRNA is obtained from a tissue, blood, urine, semen, prostatic secretion, plasma, serum, urine supernatant, urine cell pellet, or prostate cell sample.”

Col. 255, Lines 6-7
CLAIM 13 reads: “The kit of claim 12 comprising one or more composition selected from the group consisting of”

when in fact it should read:
“The kit of claim 12 comprising one or more composition selected from the group consisting of:”

Signed and Sealed this
Ninth Day of October, 2012

[Signature]
David J. Kappos
Director of the United States Patent and Trademark Office
It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

In the Claims

Col. 253 lines 66 and 67

CLAIM 7 reads:
“The kit of claim 5 comprising one or more composition selected from the group consisting of...”

when in fact it should read:

“The kit of claim 6 comprising one or more composition selected from the group consisting of...”

Signed and Sealed this  
Second Day of June, 2015

Michelle K. Lee
Director of the United States Patent and Trademark Office